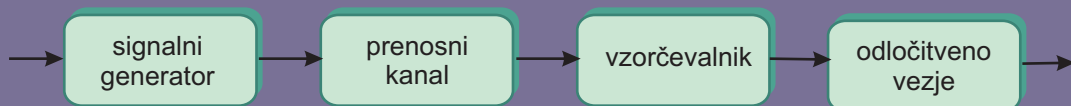
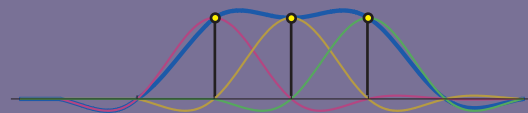
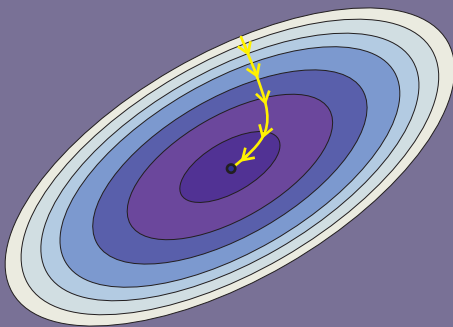
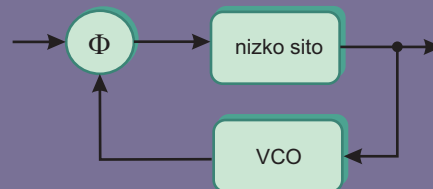
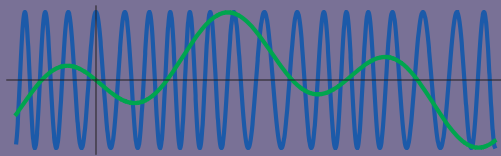


STROMAR⁵

Digitalne komunikacije

Sašo Tomažič



Ljubljana, 2012

| | | |
|-----------|--|-----------|
| 1 | Uvod | 6 |
| 2 | Signali | 10 |
| 2.1 | Periodični signali | 10 |
| 2.1.1 | Značilne vrednosti periodičnih signalov | 11 |
| 2.1.2 | Korelacija in korelacijska funkcija periodičnih signalov | 12 |
| 2.1.3 | Avtokorelacijska funkcija | 13 |
| 2.1.4 | Frekvenčni spekter periodičnih signalov | 14 |
| 2.1.5 | Amplitudni in fazni spekter periodičnih signalov | 16 |
| 2.1.6 | Močnostni spekter periodičnih signalov | 16 |
| 2.1.7 | Parsevalov izrek za periodične signale | 17 |
| 2.2 | Energijski signali | 17 |
| 2.2.1 | Enotin impulz | 18 |
| 2.2.2 | Korelacija | 18 |
| 2.2.3 | Korelacijska funkcija | 18 |
| 2.2.4 | Avtokorelacijska funkcija | 19 |
| 2.2.5 | Konvolucija | 20 |
| 2.2.6 | Frekvenčni spekter energijskih signalov | 20 |
| 2.2.7 | Amplitudni in fazni spekter energijskih signalov | 21 |
| 2.2.8 | Energijski spekter | 21 |
| 2.2.9 | Parsevalov izrek za energijske signale | 22 |
| 2.2.10 | Nekaj lastnosti Fourierovega transforma | 22 |
| 2.2.10.1 | Zakasnitev signala | 22 |
| 2.2.10.2 | Razteg in skrčenje signala | 22 |
| 2.2.10.3 | Fourierov transform realnih signalov | 23 |
| 2.2.10.4 | Fourierov transform enotinega impulza | 23 |
| 2.2.10.5 | Fourierov transform periodičnih signalov | 23 |
| 2.2.10.6 | Fourierov transform niza enotinih impulzov | 24 |
| 2.2.10.7 | Množenje s kompleksnim harmoničnim signalom | 24 |
| 2.2.10.8 | Množenje z realnim harmoničnim signalom | 25 |
| 2.2.10.9 | Fourierov transform korelacijske funkcije | 25 |
| 2.2.10.10 | Fourierov transform avtokorelacijske funkcije | 25 |
| 2.2.10.11 | Fourierov transform konvolucije | 25 |
| 2.2.10.12 | Fourierov transform produkta | 26 |

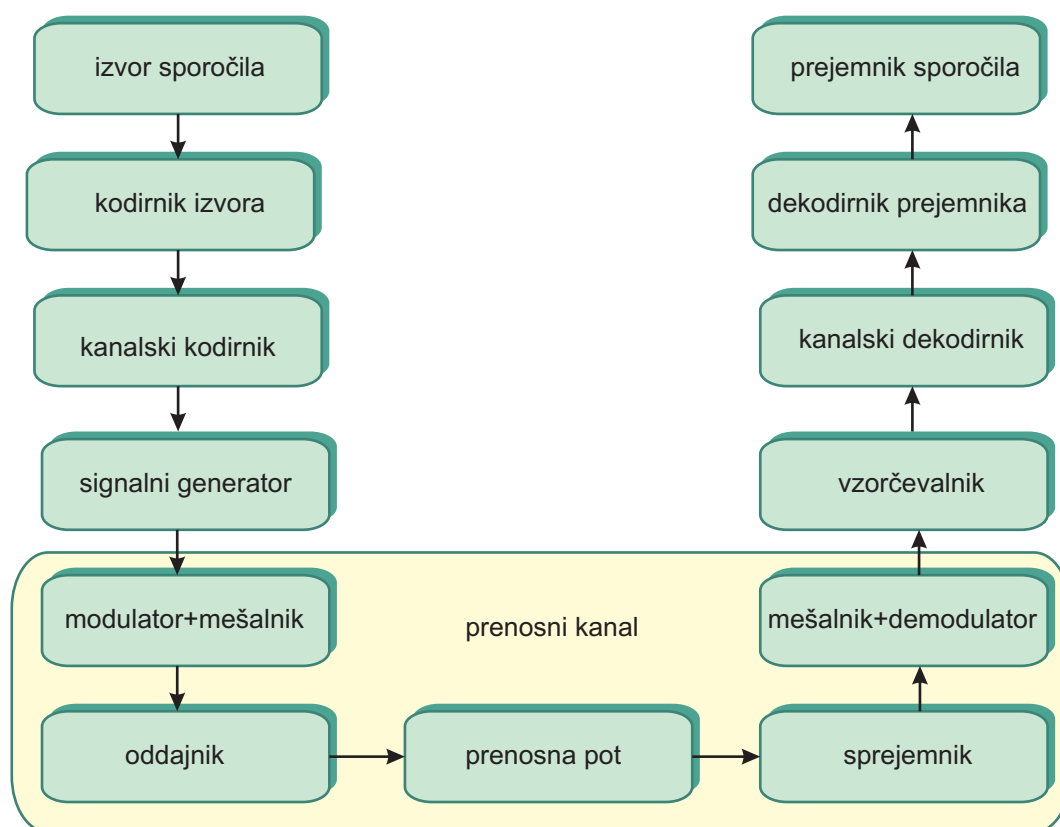
| | | |
|----------|--|-----------|
| 2.3 | Naključni signali | 26 |
| 3 | Linearni časovno nespremenljivi sistemi | 28 |
| 3.1 | Sistemska funkcija | 29 |
| 3.1.1 | Prevajalna funkcija | 29 |
| 3.2 | Kavzalni sistemi | 30 |
| 3.3 | Idealno nizko sito | 31 |
| 3.4 | Močnostni spekter pri prenosu | 31 |
| 4 | Analogen prenos | 33 |
| 4.1 | Prenos v osnovnem pasu | 33 |
| 4.2 | Prenos v višjih frekvenčnih legah | 33 |
| 4.3 | Frekvenčno mešanje | 34 |
| 4.3.1 | Mešanje navzgor | 35 |
| 4.3.2 | Mešanje navzdol | 35 |
| 4.4 | Analogni modulacijski postopki | 36 |
| 4.5 | Amplitudna modulacija | 37 |
| 4.6 | Modulacija z zadušenim nosilcem | 38 |
| 4.7 | Enobočna modulacija | 40 |
| 4.8 | Fazna modulacija | 41 |
| 4.9 | Frekvenčna modulacija | 43 |
| 4.9.1 | Demodulacija FM s fazno ujeto zanko | 44 |
| 4.9.2 | Demodulacija s FM s sukanjem faze in faznim detektorjem | 45 |
| 5 | Digitalni signali | 47 |
| 5.1 | Vzorčenje signalov | 47 |
| 5.1.1 | Teorem o vzorčenju | 48 |
| 5.2 | Spekter digitalnih signalov | 50 |
| 6 | Digitalna sita | 52 |
| 6.1 | Izvedba digitalnih sit | 52 |
| 6.2 | Adaptivna sita | 55 |
| 6.2.1 | Gradientni algoritem | 56 |
| 6.2.2 | LMS algoritem | 58 |
| 7 | Digitalen prenos v osnovnem pasu | 60 |
| 7.1 | Intersimbolna interferenca | 62 |
| 7.2 | Nyquistov kriterij za prenos brez intersimbolne interference | 63 |
| 7.2.1 | Nyquistov kriterij v časovnem prostoru | 64 |
| 7.2.2 | Nyquistov kriterij v frekvenčnem prostoru | 65 |
| 7.3 | Najožji prenosni kanal | 66 |
| 7.4 | Očesni diagram | 67 |

| | | |
|----------|---|-----------|
| 7.5 | Optimalno sprejemo sito | 69 |
| 7.6 | Izravnavna karakteristike kanala | 71 |
| 7.7 | Adaptiven izravnalnik | 72 |
| 7.8 | Adaptiven izločevalnik intersimbolne interference | 73 |
| 7.9 | Izravnalnik s prevzorčenjem | 75 |
| 7.10 | Spekter signala v osnovnem pasu | 77 |
| 8 | Digitalni modulacijski postopki | 79 |
| 8.1 | Amplitudna modulacija | 80 |
| 8.2 | Kvadratura amplitudna modulacija | 81 |
| 8.3 | Kotne modulacije | 84 |
| 8.3.1 | PSK | 85 |
| 8.3.2 | CPM | 85 |
| 8.3.3 | DPSK | 86 |
| 8.3.4 | MSK | 87 |
| 8.3.5 | GMSK | 88 |
| 8.4 | Prenos z več nosilci | 88 |
| 8.4.1 | Digitalna izvedba prenosa z več nosilci | 90 |
| 8.5 | Prenos z razširjenim spektrom | 92 |
| 8.5.1 | Frekvenčno skakanje | 92 |
| 8.5.2 | Neposredno razširjanje s kodnim zaporedjem | 93 |
| 8.5.3 | Prednosti razširjanja spektra | 95 |
| 8.5.3.1 | Zasebnost | 95 |
| 8.5.3.2 | Skritost | 95 |
| 8.5.3.3 | Odpornost proti ozkopasovnim motnjam. | 95 |
| 8.5.3.4 | Odpornost na frekvenčni presih | 96 |
| 9 | Sodostop | 97 |
| 9.1 | Frekvenčno deljen sodostop | 97 |
| 9.2 | Časovno deljen sodostop | 98 |
| 9.3 | Kodno deljen sodostop | 99 |
| 9.3.1 | Ponovna uporaba frekvenc | 100 |
| 9.3.2 | Dinamično zaseganje kanala | 101 |
| 9.3.3 | Mehko prehajanje med celicami | 101 |
| 9.3.4 | Želene lasnosti kod | 101 |
| 9.3.4.1 | Ortogonalnost in nekoreliranost | 102 |
| 9.3.4.2 | Avtokorelacijska funkcija kod | 103 |
| 9.3.5 | Povezava med frekvenčno, časovno in kodno deljenim sodostopom | 103 |
| 9.4 | Sodostop na osnovi prenosa z več nosilci | 104 |
| 9.5 | Naključno deljen sodostop | 105 |
| 9.5.1 | Protokoli ALOHA | 105 |

| | | |
|-----------|--|------------|
| 9.5.2 | Protokoli z zaznavanjem nosilca | 106 |
| 9.5.3 | Zaseganje na osnovi žetona | 107 |
| 10 | Teorija informacij | 108 |
| 10.1 | Verjetnost in informacija | 108 |
| 10.2 | Verjetnost | 109 |
| 10.3 | Odvisni simboli in pogojna verjetnost | 110 |
| 10.4 | Srednja vrednost, srednja kvadratična vrednost in varianca | 110 |
| 10.5 | Informacija | 111 |
| 10.6 | Entropija | 112 |
| 10.7 | Redundanca in irelevanca | 112 |
| 10.8 | Vzajemna informacija | 113 |
| 10.9 | Digitalni kanal | 114 |
| 10.10 | Vzajemna entropija | 114 |
| 10.11 | Kapaciteta digitalnega kanala | 115 |
| 10.12 | Binarni simetrični kanal | 116 |
| 10.12.1 | Gaussov kanal | 117 |
| 10.13 | Kapaciteta frekvenčno omejenega analognega Gaussovega kanala | 118 |
| 10.14 | Spektralna učinkovitost in Shannonova meja | 119 |
| 11 | Kanalsko kodiranje | 122 |
| 11.1 | Binarna koda | 123 |
| 11.2 | Bipolarna koda | 123 |
| 11.3 | Bifazna koda | 124 |
| 11.4 | AMI koda | 126 |
| 11.5 | HDBn kode | 127 |
| 11.6 | Binarne blokovne kode | 128 |
| 11.7 | Konovolucijske kode | 129 |
| 11.8 | Večnivojske kode | 131 |
| 11.9 | Dekodiranje na osnovi največje verjetnosti | 132 |
| 11.9.1 | Trdo odločanje | 132 |
| 11.9.2 | Mehko odločanje | 133 |
| 11.10 | Viterbi algoritem | 134 |

V poglavju je predstavljen osnovni telekomunikacijski model. Predstavljeni so osnovni pojmi, ki jih srečamo pri prenosu informacijskih signalov in so natančneje opredeljeni v naslednjih poglavjih.

Telekomunikacije so veda o prenosu sporočil na daljavo preko neke prenosne poti oziroma prenosnega medija. Celoten potek prenosa od izvora sporočila do njegovega prejemnika dobro opisuje osnovni telekomunikacijski model, ki je prikazan na sliki 1.1.



Slika 1.1 – Osnovni telekomunikacijski model. Model opisuje enosmeren prenos sporočil od njegovega izvora do njegovega prejemnika. V odvisnosti od tega, za kakšen prenos gre, so pri dejanski izvedbi lahko nekateri gradniki tega modela izpuščeni.

Osnovni telekomunikacijski model vsebuje večje število gradnikov. V odvisnosti od tega, kakšno vrsto sporočil prenašamo, kako jih prenašamo in preko česa jih prenašamo, so lahko v dejanski izvedbi prenosa nekateri gradniki izpuščeni. Ne glede na to, za kakšen prenos gre, pa so pri vsakem prenosu udeleženi *izvor sporočila*, *prenosni kanal* in *prejemnik sporočila*.

Izvor sporočila je lahko lahko človek, knjiga, računalnik, daljinski upravljalnik ali kakšna druga naprava. Od izvora sporočila je odvisna tudi oblika sporočila.

Prenosni kanal predstavlja celotno pot, po kateri se prenaša sporočilo in vključuje *oddajnik*, *prenosno pot* in *sprejemnik*.

Oddajnik je naprava, ki sporočilo pretvori v fizikalni *signal*, ki je primeren za prenos po izbrani prenosni poti. V odvisnosti od prenosne poti je ta signal lahko električen, elektromagneten, optičen ali zvočen. Oddajnik mora biti prilagojen prenosni poti.

Prenosna pot je lahko prazen prostor, zrak, kovinski vodnik, optično vlakno ali kakšna druga snov med oddajnikom in sprejemnikom, ki s svojimi fizikalnimi lastnostmi omogoča, da se po njej prenašajo signali, večinoma v obliki elektromagnetnega, lahko pa tudi zvočnega valovanja. Prenosni kanal ne izkorišča nujno vseh sposobnosti prenosne poti za prenos. Isto prenosno pot lahko izkoristimo za vzpostavitev večjega števila prenosnih kanalov. Signali se lahko pri prenosu popačijo (spremenijo obliko), hkrati pa se jim na kanalu lahko prištejejo določene motnje ali šum, kar lahko zmanjšuje kvaliteto prenosa oziroma vnaša napake v prenesena sporočila.

Sprejemnik je naprava, ki signal iz prenosne poti pretvori v obliko primerno za nadaljnjo obdelavo ali pa kar v sporočilo primerno za prejemnika sporočila. Podobno kot oddajnik mora biti tudi sprejemnik prilagojen prenosni poti.

Prejemnik sporočila je človek ali naprava, ki ji je sporočilo namenjeno. Da ima prenos sporočila smisel, mora prejemnik sporočila dobiti sporočilo v obliki, ki mu je razumljiva.

Poleg omenjenih gradnikov so lahko pri prenosu sporočil potrebni še drugi gradniki, kar pa je odvisno tako od vrste sporočil kot tudi od načina prenosa teh sporočil. Kadar želimo prenašati v višjem *frekvenčnem pasu* vključimo v prenosni kanal še *modulator*, *mešalnik* in *demodulator*.

Modulator je naprava, ki signal iz osnovnega frekvenčnega pasu prestavi v višji frekvenčni pas. Na ta način se lahko bolje prilagodimo prenosni poti oziroma lahko na isti prenosni poti ustvarimo več prenosnih kanalov.

Mešalnik je naprava, ki signal iz enega frekvenčnega pasu prestavi v drug frekvenčni pas. Prestavljanje frekvenčnega pasu po modulaciji lahko močno poenostavi izvedbo samega modulatorja, kadar želimo z istim modulatorjem preslikati v različne frekvenčne pasove oziroma z istim modulatorjem ustvariti različne prenosne kanale, kot na primer različne radijske postaje.

Demodulator je naprava, ki iz višjega frekvenčnega pasu signal preslika nazaj v osnovni pas. Postopek demodulacije je inverzen postopku modulacije.

Kadar so v prenos vključeni tudi modulator, demodulator in mešalnik, vse te elemente obravnavamo kot del prenosnega kanala.

Sporočila, ki jih želimo prenašati, so lahko v različnih oblikah. Sporočilo je lahko zapisano v signalni obliki (na primer govorni ali video signal) ali pa je zapisano s simboli (na primer črke abecede, številke ali binarni simboli v računalniškem zapisu). Kadar je sporočilo v signalni

obliki, lahko tak signal prenašamo preko prenosnega kanala neposredno, tako da vodimo signal neposredno na vhod prenosnega kanala. Tak prenos imenujemo tudi *analogen prenos*, ker je signal, ki ga prenašamo analogen izvornemu signalu.

Signal ki vsebuje sporočilo lahko pretvorimo v niz števil. Na podoben način lahko v niz števil pretvorimo tudi sporočila, ki so zapisana s poljubnimi simboli. V obeh primerih prenašamo preko prenosnega kanala niz števil (digitov), zato imenujemo tak prenos tudi *digitalen prenos*. Pri digitalnem prenosu so v prenos običajno vključeni še *kodirnik izvora*, *kanalski kodirnik*, *signalni generator*, *vzorčevalnik*, *kanalski dekodirnik* in *dekodirnik prejemnika*.

Izvorni kodirnik pretvori sporočilo iz njegove izvorne oblike v niz števil neke standardne oblike, običajno je to kar *binaren niz*. Kadar je sporočilo zapisano v signalni obliki, moramo signal najprej *vzorčiti*, to je številčno zapisati vrednosti signala v določenih časovnih intervalih. Izvorni kodirnik lahko po potrebi tudi skrči sporočilo, tako, da zmanjša potrebno *hitrost prenosa*. Izvorni kodirnik mora biti prilagojen izvoru sporočila.

Kanalski kodirnik je naprava, ki niz števil v standardni obliki pretvori v drug niz števil, ki je primernejši za prenos po kanalu. Kanalski kodirnik ima lahko vgrajene mehanizme, ki pri sprejemu omogočajo zaznavo ali pa tudi odpravo napak, do katerih lahko pride pri prenosu preko prenosnega

Signalni generator pretvori niz števil na izhodu kanalskega kodirnika v signal, ki je primeren za izbran prenosni kanal na tak način, da je pri sprejemu iz vzorcev tega signala spet mogoče dobiti nazaj originalen niz.

Vzorčevalnik je naprava, ki izmeri vrednosti signala v določenih časovnih trenutkih in jih na svoj izhod posreduje v obliki niza števil.

Kanalski dekodirnik ima inverzno funkcijo kanalskega kodirnika. Niz na svojem vhodu pretvori nazaj v standardno obliko, obenem pa lahko zazna ali tudi popravi napake, ki so nastale pri prenosu.

Kodirnik prejemnika pretvori sporočilo, ki je zapisano v standardni obliki številskega niza v obliko, ki je primerna za prejemnika sporočila. Kodirnik prejemnika ima lahko, ni pa to nujno, inverzno funkcijo kodirnika izvora. Kadar zahteva prejemnik sporočila sporočilo zapisano v obliki signala (na primer govorna komunikacija) mora kodirnik prejemnika iz vzorcev signala rekonstruirati originalni signal. Kodirnik prejemnika mora biti prilagojen prejemniku sporočila.

Opisani osnovni telekomunikacijski model opisuje zgolj preprost enosmeren prenos preko ene same samega prenosnega medija. V praksi se srečujemo s precej bolj zahtevnim prenosom. Prenos je običajno dvosmeren, prenosni medij se lahko na celotni prenosni poti večkrat spremeni, v komunikaciji lahko hkrati sodeluje več uporabnikov, več med seboj neodvisnih uporabnikov si lahko deli skupno prenosno pot in uporabniki so lahko povezani v komunikacijska omrežja.

Pri opisu osnovnega telekomunikacijskega modela smo naleteli na veliko novih pojmov, ki so zapisani v poševnem tisku. Vsi ti pojmi bodo podrobneje razloženi v nadaljevanju, kjer si bomo

podrobneje ogledali posamezne gradnike osnovnega telekomunikacijskega modela.

Pri obravnavi bomo osnovnemu telekomunikacijskemu modelu sledili od spodaj navzgor. Najprej bomo obravnavali lastnosti signalov. Razumevanje signalov je pomembno zato, da lahko razumemo, kaj se dogaja pri njihovem prenosu preko prenosne poti. Pri prenosni poti, se bomo omejili na opis lastnosti linearnih prenosnih poti, vendar se pri tem ne bomo poglobljali v fizikalno ozadje. Oddajniku in sprejemniku ne bomo posvečali večje pozornosti, ker sodi njuna obravnava bolj na področje elektronike. Veliko več pozornosti pa bomo posvetili modulacijskim in kodirnim postopkom ter drugim postopkom, ki omogočajo izboljšanje kvalitete prenosa in boljše izkoriščanje prenosnih kapacitet prenosne poti.

Kot smo že omenili, lahko določeno prenosno pot uporablja več uporabnikov hkrati, oziroma lahko preko njega ustvarimo več prenosnih kanalov, kar imenujemo sodostop do skupnega medija. To je še posebej pomembno pri brezžičnih komunikacijah, kjer si vsi uporabniki delijo skupen prenosni medij - atmosfero. Različne načine sodostopa bomo obravnavali v zadnjem poglavju.

V poglavju je definiran pojem signalov. Podana so osnovna matematična orodja za analizo periodičnih, aperiodičnih in naključnih signalov. Definirane so njihove značilne vrednosti: srednja vrednost, energija, moč in varianca ter korelacija, korelacijska funkcija in frekvenčni spekter signala.

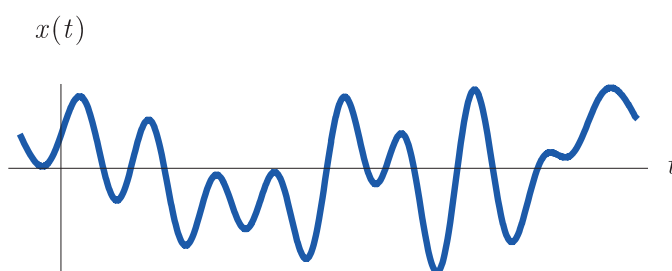
Signal je fizikalna veličina, ki se spreminja s časom. To je lahko električna napetost, električni tok, električno polje, magnetno polje, pritisk in podobno. Katera od teh veličin bo uporabljena pri prenosu sporočil je odvisno predvsem od fizikalnih lastnosti prenosne poti.

Ker se v nadaljevanju ne bomo ukvarjali s fizikalnimi lastnostmi prenosnih poti, temveč zgolj z njihovo sposobnostjo za prenos signalov, nas pri signalih ne bo zanimala njihova fizikalna narava, temveč nas bo zanimala predvsem njihova oblika, to je njihov časovni potek. Zato bomo obravnavali tako imenovane splošne signale, ki jih podajamo brez fizikalnih enot. Splošne signale pišemo kot funkcije časa t , kot na primer:

$x(t)$ za signal na vhodu nekega sistema in

$y(t)$ za signal na njegovem izhodu.

Signale pogosto predstavljamo grafično. Grafična predstavitev splošnega signala je prikazana na sliki 2.1.

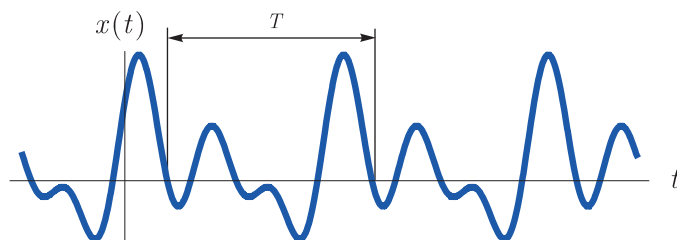


Slika 2.1 – Grafični prikaz signala

2.1 Periodični signali

Periodični signali so signali, pri katerih se začne njihov časovni potek po določenem času ponavljati. *Perioda* T signala $x(t)$ je čas ki preteče med poljubnim trenutkom in trenutkom, ko se začne potek signala ponavljati.

Grafična predstavitev periodičnega signala je prikazana na sliki 2.2.



Slika 2.2 – Grafična predstavitev periodičnega signala. Na ordinatni osi so vrednosti signala, na abscisni osi pa je čas. Po preteku periode T se začne signal ponavljati.

Periodični signal je v celoti določen s potekom znotraj ene periode, kar pomeni, da poznamo signal v celotnem časovnem področju $-\infty < t < \infty$, če poznamo njegov potek znotraj ene periode. Zato pravimo, da je periodični signal determinističen, za razliko od naključnih signalov, pri katerih ne poznamo vnaprej njihovega poteka.

Ker že vnaprej poznamo njihov potek, s periodičnimi signali ne moremo prenašati sporočil. Pri prenosu sporočil nastopajo naključni signali, saj ne vemo vnaprej kakšna sporočila se bodo prenašala preko nekega telekomunikacijskega sistema.

2.1.1 Značilne vrednosti periodičnih signalov

Značilne vrednosti so vrednosti, ki številko opredeljujejo določene lastnosti. Pri človeku so to na primer starost, teža, višina in podobno. Pomembnejše značilne vrednosti pri signalih so *perioda*, *osnovna frekvenca*, *srednja vrednost*, *srednja moč* in *varianca*

Perioda T je najmanjše pozitivno realno število, za katero velja:

$$x(t) = x(t + T) \quad \forall t \in \mathcal{R} \quad (2.1)$$

Osnovna frekvenca f_0 pove kolikokrat v časovni enoti, na primer v eni sekundi, se signal ponovi. Osnovna frekvenca je obratno sorazmerna periodi signala:

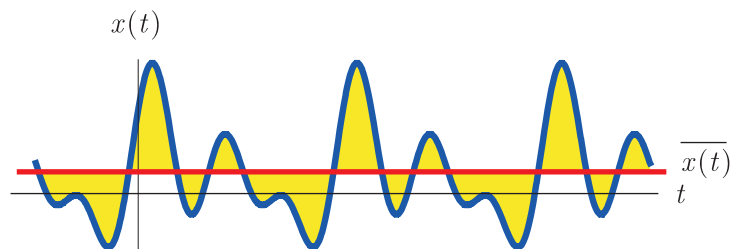
$$f_0 = \frac{1}{T} \quad (2.2)$$

Srednja vrednost ali povprečna vrednost signala je definirana tako, da je ploščina signala pod srednjo vrednostjo enaka ploščini signala nad srednjo vrednostjo, kot je to prikazano na sliki 2.3. Srednjo vrednost $\overline{x(t)}$ signala $x(t)$ s periodo T izračunamo po enačbi:

$$\overline{x(t)} = \frac{1}{T} \int_T x(t) dt \quad (2.3)$$

Oznak T na spodnji meji integrala pomeni integracijo preko ene periode signala.

Pri električnih signalih imenujemo srednjo vrednost tudi enosmerna komponenta signala.



Slika 2.3 – Srednja vrednost periodičnega signala. Ploščina signala pod srednjo vrednostjo je enaka ploščini signala nad srednjo vrednostjo.

Srednja moč ali povprečna moč. Ker je moč proporcionalna tako kvadratu napetosti kot kvadratu toka, magnetnega ali električnega polja, definiramo trenutno moč splošnih signalov z njihovim kvadratom:

$$p_x(t) = |x(t)|^2 \quad (2.4)$$

Povprečno moč periodičnega signala dobimo s povprečenjem trenutne moči preko ene periode:

$$P_x = \overline{p_x(t)} = \frac{1}{T} \int_T |x(t)|^2 dt \quad (2.5)$$

Varianca signala. Pogosto nas zanima le povprečna moč izmenične komponente signala, to je signala, ki ga dobimo, če osnovnemu signalu odštejemo njegovo srednjo vrednost. Povprečno moč izmenične komponente imenujemo tudi varianca. Pri periodičnih signalih je podana z izrazom:

$$\sigma_x^2 = \frac{1}{T} \int_T |x(t) - \overline{x(t)}|^2 dt \quad (2.6)$$

2.1.2 Korelacija in korelacijska funkcija periodičnih signalov

Korelacija signalov. Včasih želimo signale med seboj primerjati. Kot merilo podobnosti med dvema signaloma lahko uporabimo njuno korelacijo. Korelacija signalov $x(t)$ in $y(t)$, ki sta periodična s periodo T , je definirana kot srednja vrednost njunega produkta:

$$R_{xy} = \overline{x(t)y(t)} = \frac{1}{T} \int_T x(t)y(t) dt \quad (2.7)$$

Korelacijo signalov imenujemo tudi skalarni produkt signalov, ker je rezultat tega produkta skalar, to je število, in ne signal, kot pri običajnem produktu dveh signalov. Kadar je korelacija enaka 0, pravimo, da sta signala ortogonalna (prav nič podobna), po analogiji z vektorji, kjer pravimo, da sta dva vektorja ortogonalna, kadar je njun skalarni produkt enak 0.

Korelacijska funkcija. Če naredimo korelacijo med prvim signalom $x(t)$ in za τ časovno zamaknjenim drugim signalom $y(t + \tau)$ je korelacija odvisna od zamika τ , oziroma je funkcija zamika τ . Dobljeno funkcijo imenujemo korelacijska funkcija.

$$R_{xy}(\tau) = \overline{x(t)y(t + \tau)} = \frac{1}{T} \int_T x(t)y(t + \tau) dt \quad (2.8)$$

Korelacijska funkcija ima največjo vrednost pri tistem zamiku τ , pri katerem sta si signala med seboj najbolj podobna. Kadar je korelacijska funkcija dveh signalov enaka 0 pri vsakem zamiku τ , pravimo, da sta signala med seboj nekorelirana.

Primer 2.1 – Oglejmo si korelacijo in korelacijsko funkcijo signalov $x(t) = \cos(2\pi f_0 t)$ in $y(t) = \sin(2\pi f_0 t)$, kjer je f_0 osnovna frekvenca in T prerida, torej velja $f_0 = 1/T$. Njuna korelacija je enaka:

$$R_{xy} = \frac{1}{T} \int_T \cos(2\pi f_0 t) \sin(2\pi f_0 t) dt = \frac{1}{2T} \int_T \sin(4\pi f_0 t) dt = 0 \quad (2.9)$$

Drugi integral je očitno enak nič, saj integriramo sinus čez dve celi periodi. Ker je skalarni produkt enak 0 sta signala $\cos(2\pi f_0 t)$ in $\sin(2\pi f_0 t)$ med seboj ortogonalna, kar pa še ne pomeni, da si med seboj nista prav nič podobna, da nista korelirana. Njuna korelacijska funkcija je namreč enaka:

$$R_{xy}(\tau) = \frac{1}{T} \int_T \cos(2\pi f_0 t) \sin(2\pi f_0(t + \tau)) dt = \frac{1}{2} \sin(2\pi f_0 \tau) \quad (2.10)$$

Korelacijska funkcija ima maksimalno vrednost pri $\tau = T/4$, to je takrat, ko postane sinus zaradi časovnega premika enak kosinusu.

2.1.3 Avtokorelacijska funkcija

Če v korelacijski funkciji (2.8) nadomestimo drugi signal, to je $y(t)$, s prvim signalom, to je $x(t)$, dobimo avtokorelacijsko funkcijo:

$$R_{xx}(\tau) = \overline{x(t)x(t + \tau)} = \frac{1}{T} \int_T x(t)x(t + \tau) dt \quad (2.11)$$

Avtokorelacijska funkcija je torej korelacijska funkcija, ki primerja signal s z njegovo premaknjeno različico. Ker je avtokorelacijska funkcija pomembna pri analizi signalov, si oglejmo nekaj njenih lastnosti:

1. Avtokorelacijska funkcija periodičnega signala je periodična. Perioda avtokorelacijske funkcije je enaka periodi signala:

$$R_{xx}(\tau) = R_{xx}(\tau + T) \quad (2.12)$$

2. Avtokorelacijska funkcija je soda funkcija:

$$R_{xx}(\tau) = R_{xx}(-\tau) \quad (2.13)$$

3. Avtokorelacijska funkcija je maksimalna pri vrednosti $\tau = 0$:

$$R_{xx}(0) \geq R_{xx}(\tau) \quad (2.14)$$

4. Vrednost avtokorelacijske funkcije periodičnega signala pri $\tau = 0$ je enaka povprečni moči signala:

$$R_{xx}(0) = \frac{1}{T} \int_T x^2(t) dt = \overline{x^2(t)} \quad (2.15)$$

2.1.4 Frekvenčni spekter periodičnih signalov

Vsak periodičen signal s končno močjo lahko razvijemo v kompleksno Fourierovo vrsto:

$$x(t) = \sum_{k=-\infty}^{\infty} X_k e^{j2\pi k f_0 t} \quad (2.16)$$

Zgornja enačba pove, da lahko vsak periodičen signal končne moči zapišemo kot vsoto posameznih spektralnih komponent, to je kompleksnih harmoničnih signalov frekvenc, ki so mnogokratniki osnovne frekvence $e^{j2\pi k f_0}$. Te komponente imenujemo tudi harmonske komponente signala in jih izračunamo po enačbi:

$$X_k = \frac{1}{T} \int_0^T x(t) e^{j2\pi k f_0 t} dt \quad (2.17)$$

Pri realnih signalih, $x(t) \in \Re$, velja, da so vrednosti koeficientov pri negativnih k enaki konjugirano kompleksnim vrednostim koeficientov pri pozitivnih k :

$$X_{-k} = X_k^* \quad (2.18)$$

Posamezno harmonsko komponento realnega signala dobimo s seštevanjem kompleksnih komponent pri pozitivni in pripadajoči negativni frekvenci:

$$x_k(t) = X_{-k} e^{-j2\pi k f_0 t} + X_k e^{j2\pi k f_0 t} \quad (2.19)$$

kjer smo z $x_k(t)$ označili harmonsko komponento signala pri frekvenci $k f_0$. Če sedaj v zgornji enačbi upoštevamo enačbo (2.18) in dejstvo da velja:

$$e^{j2\pi k f_0 t} = \cos(2\pi k f_0 t) + j \sin(2\pi k f_0 t)$$

dobimo po krajši izpeljavi:

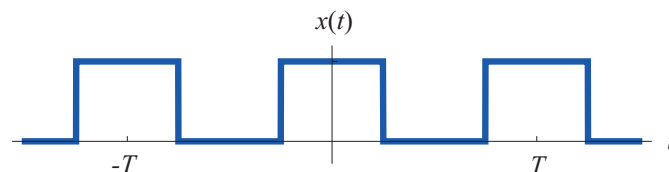
$$x_k(t) = 2 \Re(X_k) \cos(2\pi k f_0 t) + 2 \Im(X_k) \sin(2\pi k f_0 t) \quad (2.20)$$

Vidimo, da je posamezna harmonska komponenta sestavljena iz kosinusne in sinusne komponente. Realni del koeficientov X_k predstavlja polovico amplitude kosinusnega signala, njihov imaginarni del pa polovico amplitude sinusnega signala pri ustrezni frekvenci.

Pri realnih signalih je spekter pri negativnih frekvencah samo zrcalna slika spektra pri pozitivnih frekvencah. Posamezne harmonske komponente dobimo s seštevanjem pripadajočih kompleksnih komponent pri pozitivni in pripadajoči negativni frekvenci. Negativne frekvence zato nimajo fizikalnega pomena.

Primer 2.2 – Pravokoten periodičen signal. Naj bo $x(t)$ pravokoten periodičen signal, ki je prikazan na sliki (2.4). Po enačbi (2.17) lahko izračunamo koeficiente X_k :

$$X_k = \frac{1}{T} \int_{-T/4}^{T/4} e^{j2\pi k f_0 t} dt = \frac{\sin(k \frac{\pi}{2})}{k \frac{\pi}{2}} \quad (2.21)$$



Slika 2.4 – Pravokoten periodičen signal

Ker je signal $x(t)$ soda funkcija (to je simetrična okrog ordinatne osi), jo je mogoče dobiti s seštevanjem samih kosinuso. Zato so koeficienti realni in nastopajo pri razvoju v Fourierovo vrsto samo kosinusi z amplitudami:

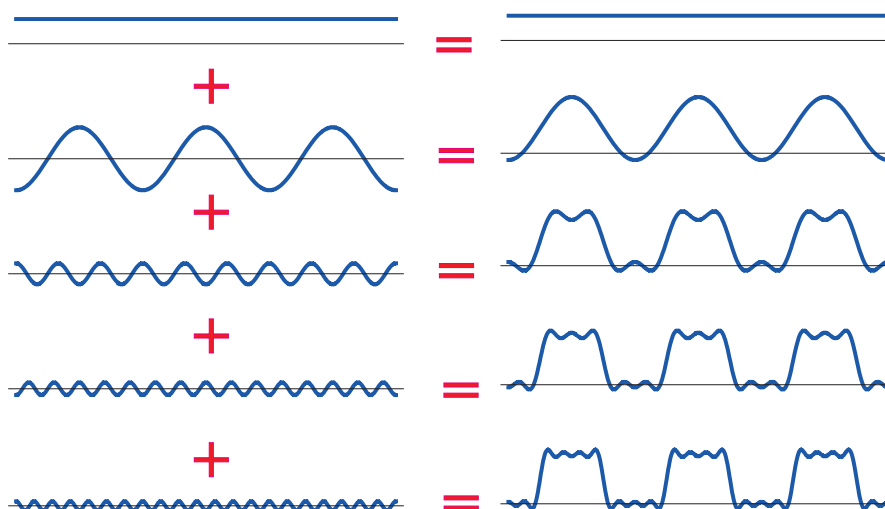
$$a_k = X_k + X_k^* = 2X_k$$

Prvih nekaj koeficientov a_k različnih od 0 je podanih v tabeli 2.1

| k | a_k |
|-----|-------------|
| 0 | $1/2$ |
| 1 | $2/\pi$ |
| 3 | $-2/(3\pi)$ |
| 5 | $2/(5\pi)$ |
| 7 | $-2/(7\pi)$ |

Tabela 2.1 – Amplitude kosinusnih harmonskih komponent pri razvoju sodega pravokotnega signala v Fourierovo vrsto.

Kako se s seštevanjem posameznih harmonskih komponent približujemo pravokotnemu signalu je prikazano na sliki 2.5



Slika 2.5 – Seštevanje harmonskih komponent pravokotnega signala

2.1.5 Amplitudni in fazni spekter periodičnih signalov

Koeficienti X_k pri razvoju signala so v splošnem kompleksni

$$X_k = \Re(X_k) + j\Im(X_k) \quad (2.22)$$

in jih lahko pišemo v polarni obliki:

$$X_k = A_k e^{j\Phi_k} \quad (2.23)$$

Koeficiente A_k imenujemo amplitudni spekter periodičnega signala. Velja:

$$A_{X_k} = |X_k| = \sqrt{\Re^2(X_k) + \Im^2(X_k)} \quad (2.24)$$

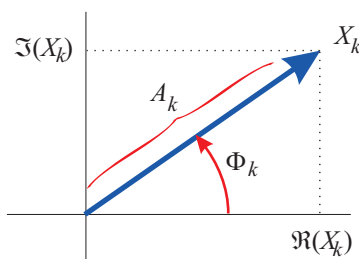
Koeficiente Φ_k imenujemo fazni spekter periodičnega signala. Velja:

$$\Phi_{X_k} = \arg(X_k) \quad (2.25)$$

Funkcija $\arg[X_k]$ predstavlja kot med realno osjo in kazalcem X_k v kompleksni ravnini, kot je to prikazano na sliki ???. Definira jo naslednji izraz:

$$\arg(X_k) = \begin{cases} \arctan\left(\frac{\Im(X_k)}{\Re(X_k)}\right) & ; \Re(X_k) \geq 0 \\ \arctan\left(\frac{\Im(X_k)}{\Re(X_k)}\right) + \pi & ; \Re(X_k) < 0 \end{cases} \quad (2.26)$$

Grafični prikaz polarne zapisa je podan na sliki 2.6.



Slika 2.6 – Kazalec komponente kompleksnega spektra. Dolžina kazalca predstavlja komponento amplitudnega spektra, kot med kazalcem in realno osjo pa komponento faznega spektra.

2.1.6 Močnostni spekter periodičnih signalov

Po preprostem izračunu ugotovimo, da je moč posamezne spektralne komponente, kot je podana z enačbo 2.20, enaka:

$$\overline{x_k^2(t)} = \int_T x_k^2(t) dt = 2|X_k|^2 = |X_k|^2 + |X_{-k}|^2 \quad (2.27)$$

Polovico moči posamezne realne spektralne komponente prispeva kompleksna komponenta pri negativni frekvenci in polovico komponenta pri pozitivni frekvenci. Močnostni spekter S_{X_k} periodičnega signala zato definiramo s kvadratom absolutnih vrednosti kompleksnega spektra:

$$S_{X_k} = |X_k|^2 \quad (2.28)$$

2.1.7 Parsevalov izrek za periodične signale

Moč periodičnega signala je enaka vsoti moči posameznih harmonskih komponent:

$$\overline{x^2(t)} = \frac{1}{T} \int_0^T x^2(t) dt = \sum_{k=-\infty}^{\infty} |X_k|^2 = \sum_{k=-\infty}^{\infty} S_{Xk} \quad (2.29)$$

V splošnem moč vsote signalov ni enaka vsoti moči posameznih signalov. To, da se moči posameznih harmonskih komponent seštevajo, je posledica dejstva, da so te komponente med seboj ortogonalne. To lahko v določenih primerih izredno poenostavi računanje moči signalov.

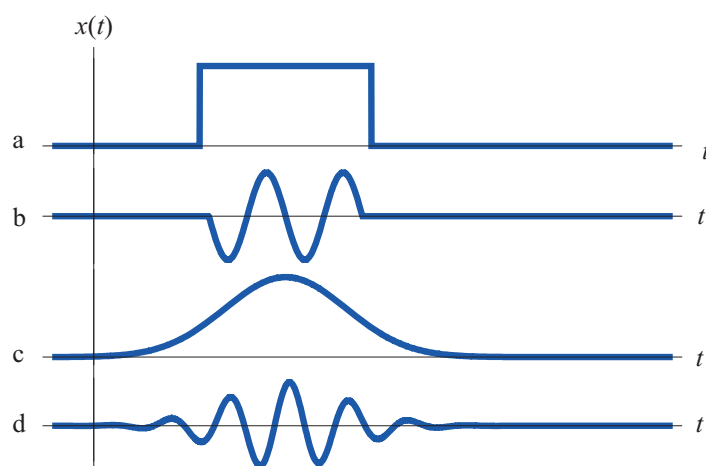
2.2 Energijski signali

Energijski signali so signali, ki imajo končno energijo.

Energija signala je značilna vrednost signala. Energija E_x signala $x(t)$ je enaka integralu trenutne moči $x^2(t)$ in je podana z izrazom:

$$E_x = \int_{-\infty}^{\infty} |x(t)|^2 dt \quad (2.30)$$

Ker imajo energijski signali končno energijo imajo tudi omejen čas trajanja. Signale, ki imajo relativno kratek čas trajanja, imenujemo tudi impulzi. Primeri različnih impulzov so prikazani na sliki 2.7.



Slika 2.7 – Primeri impulzov. a - pravokoten impulz, b - moduliran pravokoten impulz, c - Gaussov impulz, d - moduliran Gaussov impulz

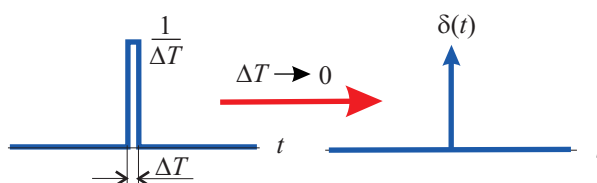
2.2.1 Enotin impulz

Enotin impulz $\delta(t)$ je neskončno ozek impulz z neskončno amplitudo, tako da je ploščina pod njim enaka 1:

$$\int_{-\epsilon}^{\epsilon} \delta(t) dt = 1 \quad (2.31)$$

kjer je ϵ poljubno majhno pozitivno število.

Enotin impulz imenujemo tudi Diracova funkcija ali δ -impulz. Grafično ga predstavimo z navpično puščico (slika 2.8, desno).



Slika 2.8 – Približek enotnega impulza. Z zmanjševanjem časa ΔT se vedno bolj približujemo enotnemu impulzu.

Čeprav enotin impulz v naravi ne nastopa, ima zaradi svojih lastnosti pomembno vlogo pri analizi signalov. V naravi se lahko enotnemu impulzu približamo s kratkim pravokotnim impulzom trajanja ΔT in amplitude $1/\Delta T$, kot je to prikazano na sliki 2.8.

Pomembna lastnost enotnega impulza je njegova lastnost vzorčenja:

$$\int_{-\infty}^{\infty} x(t)\delta(t - t_0) dt = x(t_0) \quad (2.32)$$

Integral poljubnega signala pomnoženega z za t_0 zakasnjanim enotnim impulzom je enaka vrednosti signala ob času t_0 .

2.2.2 Korelacija

Korelacija ali skalarni produkt signalov $x(t)$ in $y(t)$ je definiran kot integral njunega produkta:

$$r_{xy} = \int_{-\infty}^{\infty} x(t)y(t) dt \quad (2.33)$$

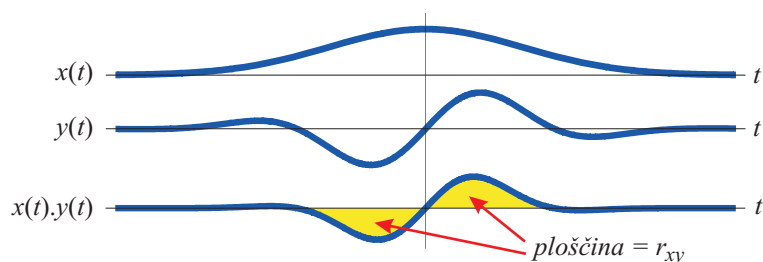
Korelacija je merilo podobnosti dveh signalov. Kadar je korelacija enaka 0 pravimo, da sta signala ortogonalna. Grafična predstavitev korelacijske funkcije je prikazana na sliki 2.9.

2.2.3 Korelacijska funkcija

Korelacijska funkcija je funkcija, ki jo dobimo, če naredimo korelacijo dveh časovno zamaknjenih signalov:

$$r_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)y(t + \tau) dt \quad (2.34)$$

Korelacijska funkcija je funkcija časovnega zamika τ . Kadar je korelacijska funkcija pri vseh časovnih zamikih enaka 0, pravimo, da sta signala nekorelirana.



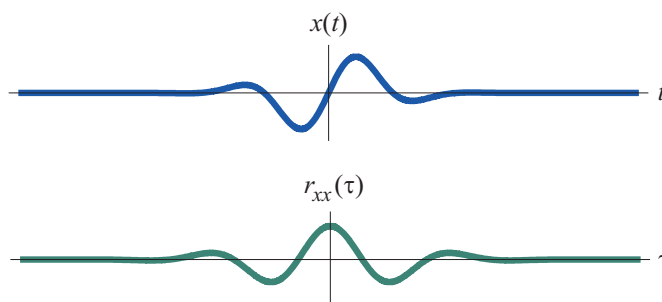
Slika 2.9 – Korelacija je enaka ploščini pod krivuljo produkta obeh signalov. Kjer imata oba signala enak predznak je ploščina pozitivna, kjer pa imata nasproten predznak je negativna. Kadar je negativna ploščina enaka pozitivni, je korelacija enaka 0 in signala sta ortogonalna.

2.2.4 Avtokorelacijska funkcija

Če v izrazu (2.34) signal $y(t)$ nadomestimo s signalom $x(t)$ dobimo avtokorelacijsko funkcijo:

$$r_{xx}(\tau) = \int_{-\infty}^{\infty} x(t)x(t+\tau) dt \quad (2.35)$$

Avtokorelacijska funkcija je funkcija časovnega zamika τ . Pri avtokorelacijski funkciji primerjamo signal z njegovo časovno zamaknjeno različico. Primer avtokorelacijske funkcije signala je prikazan na sliki 2.10



Slika 2.10 – Avtokorelacijska funkcija je funkcija časovnega zamika τ . Je soda funkcija in ima maksimum pri $\tau = 0$

Oglejmo si nekaj lastnosti avtokorelacijske funkcije:

1. Avtokorelacijska funkcija je soda funkcija:

$$r_{xx}(\tau) = r_{xx}(-\tau) \quad (2.36)$$

2. Avtokorelacijska funkcija je maksimalna pri vrednosti $\tau = 0$:

$$r_{xx}(0) \geq r_{xx}(\tau) \quad (2.37)$$

3. Vrednost avtokorelacijske funkcije pri $\tau = 0$ je enaka energiji signala:

$$r_{xx}(0) = \int_{-\infty}^{\infty} x^2(t) dt = E_x \quad (2.38)$$

2.2.5 Konvolucija

Konvolucija signalov $x(t)$ in $y(t)$ je korelacija prvega signala s po času premaknjeno in obrnjeno različico drugega signala, pri čemer zamenjamo pomen spremenljivk t in τ , tako, da ostane korelacija funkcija časa. Korelacija predstavlja nekakšen produkt dveh signalov, zato operacijo konvolucije označujem z zvezdico:

$$x(t) * y(t) == \int_{-\infty}^{\infty} x(\tau)x(t - \tau) d\tau \quad (2.39)$$

Konvolucija je komutativna operacija, tako da velja:

$$x(t) * y(t) = y(t) * x(t) \quad (2.40)$$

Iz lastnosti vzorčenja, ki jo ima enotin impulz, neposredno sledi:

$$x(t) * \delta(t) = x(t) \quad (2.41)$$

Enotin predstavlja torej enoto za operacijo konvolucije. Od tod izvira tudi njegovo ime enotin impulz.

2.2.6 Frekvenčni spekter energijskih signalov

Frekvenčni spekter energijskih signalov računamo s pomočjo Fourierovega transformata:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt \quad (2.42)$$

Spekter signala pri frekvenci f je pravzaprav korelacija tega signala s kompleksnim harmoničnim signalom $e^{-j2\pi ft}$, torej $X(f)$ pove, koliko sta si ta dva signala podobna.

Za razliko od spektra periodičnih signalov, v katerem nastopajo samo mnogokratniki osnovne frekvence signala (diskreten spekter), lahko v spektru energijskih signalov nastopajo spektralne komponente zvezno pri vseh frekvencah f (zvezen spekter).

Podobno kot pri periodičnih signalih, dobimo signal s seštevanjem vseh spektralnih komponent. Ker je spekter zvezen nastopa v enačbi namesto vsote integral:

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft} df \quad (2.43)$$

Zgornjo enačbo imenujemo tudi inverzni Fourierov transform.

Funkciji $x(t)$ in $X(f)$ imenujemo Fourierov par, kar označujemo z dvojno puščico:

$$x(t) \leftrightarrow X(\omega) \quad (2.44)$$

V splošnem se bomo držali pravila, da signale označujemo z malimi črkami, za spekter pa uporabimo iste velike črke, tako da že iz samih oznak lahko razberemo, da gre za Fourierov par.

Fourierov transform označujemo krajše tudi z \mathcal{F} :

$$X(\omega) = \mathcal{F}[x(t)] \quad (2.45)$$

inverzni Fourierov transform pa z \mathcal{F}^{-1} :

$$x(t) = \mathcal{F}^{-1}[X(f)] \quad (2.46)$$

2.2.7 Amplitudni in fazni spekter energijskih signalov

Podobno kot pri periodičnih signalih, je amplitudni spekter $A_x(f)$ signala $x(t)$ enak absolutni vrednosti njegovega kompleksnega spektra $X(f)$:

$$A_x(f) = |X(f)| \quad (2.47)$$

fazni spekter $\Phi_x(f)$ pa argumentu kompleksnega spektra:

$$\Phi_x(f) = \arg(X(f)) \quad (2.48)$$

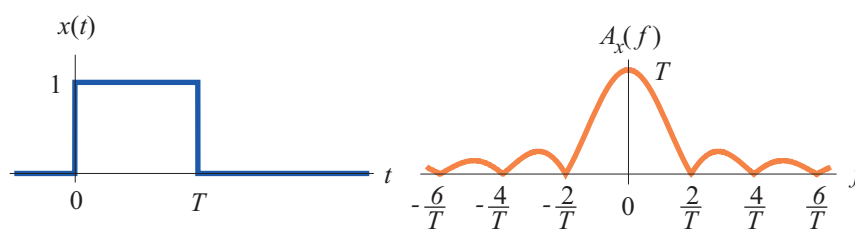
Primer 2.3 – Amplitudni spekter pravokotnega impulza. Kot primer si oglejmo izračun amplitudnega spektra pravokotnega impulza dolžine T , ki je prikazan na levi strani slike 2.11. Najprej izračunamo njegov kompleksni spekter:

$$\begin{aligned} X(f) &= \int_0^T e^{-j2\pi ft} dt = \\ &= \frac{e^{-j2\pi fT} - 1}{-j2\pi f} = \\ &= T \cdot e^{-j\pi fT} \cdot \frac{e^{j\pi fT} - e^{-j\pi fT}}{j2\pi fT} = \\ &= T \cdot e^{-j\pi fT} \cdot \frac{\sin(\pi fT)}{\pi fT} \end{aligned} \quad (2.49)$$

in nato še njegovo absolutno vrednost, da dobimo amplitudni spekter:

$$A_x(f) = |X(f)| = T \cdot \left| \frac{\sin(\pi fT)}{\pi fT} \right| \quad (2.50)$$

Amplitudni spekter pravokotnega impulza je prikazan na desni strani slike 2.11.



Slika 2.11 – Amplitudni spekter pravokotnega impulza. Spekter ima pri frekvenci 0 vrednost T in prvo ničlo pri frekvenci $2/T$. Če manjšamo T , se amplituda zmanjšuje, frekvenčno področje pa se širi.

2.2.8 Energijski spekter

Pri energijskih signalih govorimo o energijskem spektru in ne o močnostnem spektru, kot pri periodičnih signalih, ki imajo neskončno energijo in omejeno moč. Energijski spekter $E_x(f)$ signala $x(t)$ je enak kvadratu absolutne vrednosti njegovega spektra:

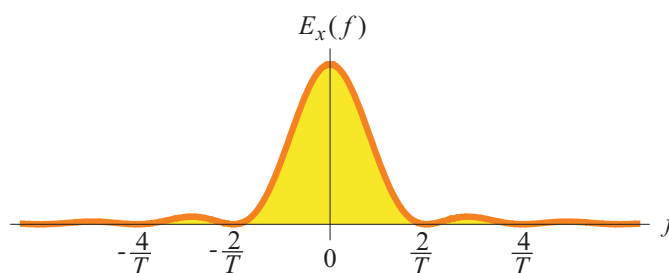
$$E_x(f) = |X(f)|^2 \quad (2.51)$$

2.2.9 Parsevalov izrek za energijske signale

Energija signala je enaka integralu energijskega spektra:

$$E_x = \int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} E_x(f) df \quad (2.52)$$

Energijski spekter pravokotnega impulza je prikazan na sliki 2.12. Energija signala je enaka ploščini pod krivuljo energijskega spektra.



Slika 2.12 – Energijski spekter pravokotnega impulza dolžine T . Rumeno obarvana ploščina pod krivuljo je enaka energiji impulza. Večina energije impulza je v frekvenčnem področju med frekvencami 0 in $2/T$

2.2.10 Nekaj lastnosti Fourierovega transformata

Ker je Fourierov transform zelo pomembno orodje pri analizi signalov, si oglejmo nekaj njegovih pomembnejših lastnosti. Pri navajanju vseh lastnosti predstavljata $x(t)$ in $X(f)$ Fourierov par.

2.2.10.1 Zakasnitev signala

Signal $x(t)$ zakasnimo za čas t_0 .

$$x(t - t_0) \leftrightarrow e^{-j2\pi t_0 f} X(f) \quad (2.53)$$

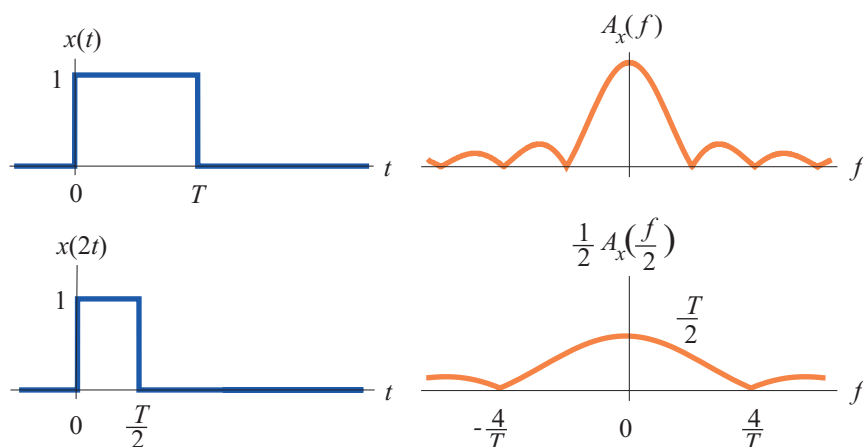
Pri zakasnitvi ostane absolutna vrednost Fourierovega transformata (amplitudni spekter) nespremenjena, fazi pa se prišteje faza $-2\pi t_0 f$, ki se linearno spreminja s frekvenco f .

2.2.10.2 Razteg in skrčenje signala

Čas t množimo z realno z realno konstanto a . Če je $a < 1$ se signal raztegne, če pa je $a > 1$ se signal skrči po časovni osi. Za spekter takega signala velja:

$$x(at) \leftrightarrow \frac{1}{|a|} X\left(\frac{f}{a}\right) \quad (2.54)$$

Ker se v spektru frekvenca f deli z a , pomeni razteg v časovnem prostoru skrčenje v frekvenčnem prostoru in obratno. Iz tega neposredno sledi, da imajo kratki impulzi širok spekter, dolgi impulzi pa ozek spekter, kot je to prikazano na sliki 2.13



Slika 2.13 – Časovno skrčenje signala razširi njegov spekter. Ožji impulz ima zato širši spekter.

2.2.10.3 Fourierov transform realnih signalov

Če je signal $x(t)$ realen, potem je njegov spekter pri negativnih frekvencah konjugirano kompleksna zrcalna slika spektra pri pozitivnih frekvencah:

$$X(-f) = X^*(f) \quad (2.55)$$

zato je amplitudni spekter soda funkcija frekvence, fazni spekter pa le liha funkcija frekvence.

Podobno kot pri periodičnih signalih tudi pri energijskih signalih spekter pri negativnih frekvencah nima fizikalnega pomena. Spekter pri negativnih frekvencah je zgolj zrcalna slika spektra pri pozitivnih frekvencah.

2.2.10.4 Fourierov transform enotnega impulza

Naj bo signal $x(t)$ enak enotnemu impulzu. Iz lastnosti o vzorčenju neposredno sledi:

$$X(f) = \int_{-\infty}^{\infty} \delta(t) e^{-j2\pi ft} dt = e^0 = 1 \quad (2.56)$$

Spekter enotnega impulza je konstanten oziroma raven, kar pomeni, da vsebuje vse frekvence. Po analogiji z belo svetlobo, ki vsebuje vse frekvence, imenujemo tak spekter tudi bel spekter.

2.2.10.5 Fourierov transform periodičnih signalov

Čeprav smo spekter harmoničnih signalov računali z razvojem v Fourierovo vrsto, pa ga lahko izrazimo tudi s pomočjo Fourierovega transformata. Velja:

$$X(f) = \sum_{k=-\infty}^{\infty} X_k \delta(f - kf_0) \quad (2.57)$$

kjer je f_0 osnovna frekvenca signala, X_k pa so koeficienti, ki jih dobimo z razvojem v Fourierovo vrsto. Posamezne harmonske komponente so tu predstavljene z enotnimi impulzi pri ustrezni frekvenci. Funkcija $X(f)$ v resnici predstavlja gostoto amplitudnega spektra. Ker so v periodičnih signalih prisotne harmonske komponente s točno določenih frekvenc, je gostota spektra pri teh frekvencah neskončna, kar ponazarjajo enotni impulzi.

2.2.10.6 Fourierov transform niza enotnih impulzov

Niz enotnih impulzov, ki so razmaknjeni za čas T je periodičen signal z osnovno frekvenco $f_0 = 1/T$ in ga lahko zapišemo v obliki:

$$x(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT) \quad (2.58)$$

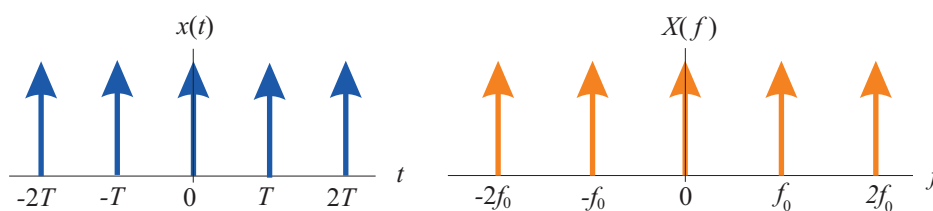
Če izračunamo spekter tega signala po enačbi (2.17) dobimo:

$$X_k = \frac{1}{T} \quad (2.59)$$

Signal torej vsebuje vse harmonske komponente z enako amplitudo $1/T$. V skladu z izrazom za Fourierov transform periodičnih signalov (2.57) je torej Fourierov transform niza enotnih impulzov enak:

$$X(f) = \frac{1}{T} \sum_{k=-\infty}^{\infty} \delta(f - kf_0) \quad (2.60)$$

Niz enotnih impulzov s časovnim razmikom T se s Fourierovim transformom preslika v niz enotnih impulzov s frekvenčnim razmikom f_0 , kot je to prikazano na sliki 2.14.



Slika 2.14 – Fourierov transform niza enotnih impulzov je niz enotnih impulzov

2.2.10.7 Množenje s kompleksnim harmoničnim signalom

Če signal množimo s kompleksnim harmoničnim signalom s frekvenco f_0 se njegov spekter premakne v desno za f_0 :

$$x(t)e^{j2\pi f_0 t} \leftrightarrow X(f - f_0) \quad (2.61)$$

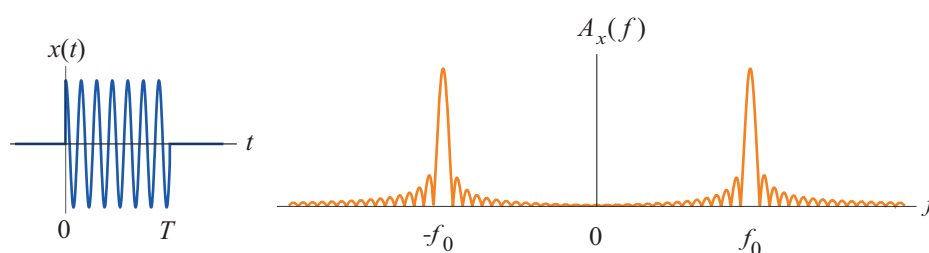
Ker signal po množenju s kompleksnim harmoničnim signalom ni več realen, se tudi spekter ne zrcali več preko ordinatne osi.

2.2.10.8 Množenje z realnim harmoničnim signalom

Če signal $x(t)$ množimo z realnim harmoničnim signalom s frekvenco f_0 , se njegov premakne levo in desno za frekvenco f_0 , tako da se ohrani simetrija spektra preko ordinatne osi:

$$x(t) \cos(2\pi f_0 t) \leftrightarrow \frac{1}{2} (X(f - f_0) + X(f + f_0)) \quad (2.62)$$

Na sliki 2.15 je prikazan amplitudni spekter impulza, ki ga dobimo, če pravokotni impulz dolžine T množimo s kosinusom frekvence $f - 0$.



Slika 2.15 – Množenje pravokotnega signala s $\cos(2\pi f_0 t)$. Spekter pravokotnega signala se premakne levo in desno za frekvenco f_0 .

2.2.10.9 Fourierov transform korelacijske funkcije

Če sta $X(\omega)$ in $Y(\omega)$ Fourierova transformata funkcij $x(t)$ in $y(t)$ velja:

$$r_{xy}(\tau) \leftrightarrow X^*(\omega)Y(\omega) \quad (2.63)$$

2.2.10.10 Fourierov transform avtokorelacijske funkcije

Iz (2.63) neposredno sledi:

$$r_{xx}(t) \leftrightarrow |X(f)|^2 = E_x(f) \quad (2.64)$$

kar pomeni, da je gostota energijskega spektra aperiodičnega signala enaka Fourierovemu transformu njegove avtokorelacijske funkcije.

2.2.10.11 Fourierov transform konvolucije

Fourierov transform konvolucije signalov $x(t)$ in $y(t)$ je enak produktu Fourierovih transformov teh signalov $X(f)$ in $Y(f)$:

$$x(t) * y(t) \leftrightarrow X(f) \cdot Y(f) \quad (2.65)$$

Rečemo lahko, da se konvolucija s Fourierovim transformom pretvori v produkt.

2.2.10.12 Fourierov transform produkta

Fourierov transform produkta signalov $x(t)$ in $y(t)$ je enak konvoluciji Fourierovih transformov teh signalov $X(f)$ in $Y(f)$:

$$x(t) \cdot y(t) \leftrightarrow X(f) * Y(f) \quad (2.66)$$

Fourierov transform transformira produkt v konvolucijo.

Čeprav običajno govorimo, da Fourierov transform preslika signal iz časovnega prostora v frekvenčni prostor, se moramo zavedati, da je Fourierov transform samo orodje za analizo signala. Spekter signala je samo ena od lastnosti signala, podobno kot energija signala. Ko izračunamo energijo signala, se signal v ničemer ne spremeni. Na enak način signal ostane nespremenjen, ko izračunamo njegov spekter.

2.3 Naključni signali

Naključni signali so signali, za katere ne poznamo natančno njihovega poteka, lahko pa poznamo nekaj statističnih lastnosti. Tipično so to signali, ki jih prenašamo. Za signale, ki jih prenašamo, namreč vnaprej ne poznamo poteka. Če bi ga poznali vnaprej, potem jih ne bi bilo potrebno prenašati. Primeri takih signalov so govorni signal, avdio signal, video signal, podatkovni signal in podobno.

Ker predpostavljamo, da časovno niso omejeni, oziroma v praksi lahko trajajo relativno zelo dolgo z ozirom na hitrost s katero se spreminjajo, njihova energija ni omejena, zato Fourierov transform takih signalov ne obstaja. Povprečna moč takega signala je dana z izrazom:

$$P_x = \overline{|x(t)|^2} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(t)|^2 dt \quad (2.67)$$

Predpostavimo, da je moč signala končna. Za tak signal definiramo avtokorelacijsko funkcijo:

$$R_{xx}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(\tau)x(t + \tau) d\tau \quad (2.68)$$

in močnostni spekter:

$$S_x(f) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-j2\pi\tau} d\tau \quad (2.69)$$

Avtokorelacijska funkcija govori o tem, kako hitro se signal spreminja. Če signal sam sebi ni več podoben, ko ga malo zamaknemo po času, je avtokorelacijska funkcija ozka. Iz lastnosti Fourierovega transformata vemo, da je potem spekter signala širok. Skrajni primer dobimo, ko je avtokorelacijska funkcija enaka enotnemu impulzu. Takrat je močnostni spekter neskončno širok in raven. Pravimo, da je spekter bel.

Močnostni spekter lahko izrazimo tudi direktno s časovnim signalom kot:

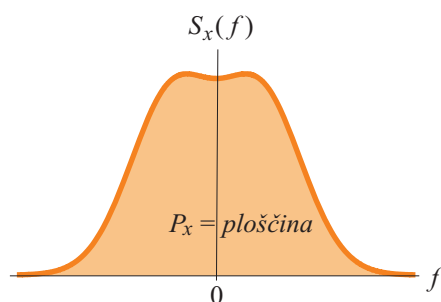
$$S_x(f) = \lim_{T \rightarrow \infty} \frac{1}{T} \left| \int_{-T/2}^{T/2} x(t) e^{-j2\pi ft} dt \right|^2 \quad (2.70)$$

vendar to lahko storimo šele, ko signal poznamo v celoti. Čas T , ki gre v zgornji enačbi proti neskončnosti, pa je v praksi omejen na trajanje signala.

Parsevalov izrek za močnostne signale govori o moči signala in pravi, da je moč signala enaka integralu močnostnega spektra

$$P_x = \overline{|x(t)|^2} = \int_{-\infty}^{\infty} S_x(f) df \quad (2.71)$$

oziroma ploščini pod krivuljo močnostnega spektra, kot je to prikazano na sliki 2.16.



Slika 2.16 – Parsevalov izrek pravi, da je moč signala P_x enaka ploščini pod krivuljo močnostnega spektra $S_x(f)$.

Signali, ki jih prenašamo v telekomunikacijah so vedno naključni. Če bi vnaprej poznali njihov potek, prenos ne bi imel smisla. Pri telekomunikacijah gre namreč za prenos informacije in, kot bomo spoznali v nadaljevanju, lahko samo naključni signali vsebujejo informacijo.

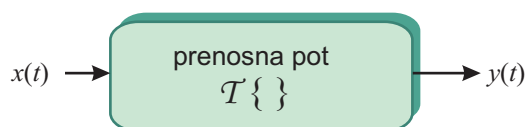
Linearni časovno nespremenljivi sistemi

Poglavje obravnava lastnosti linearnih časovno nespremenljivih sistemov. Večino prenosnih poti lahko namreč obravnavamo kot linearne časovno nespremenljive sisteme. Frekvenčna sita ali filtri, ki prepuščajo določene frekvence, druge pa slabijo, ravno tako spadajo v skupino linearnih časovno nespremenljivih sistemov.

Signale prenašamo kot valovanje preko različnih prenosnih poti. Večinoma gre za elektromagnetno valovanje, redkeje pa tudi za zvočno valovanje. Pri prenosu preko prenosne poti signal spremeni svojo obliko. Pravimo, da se signal popači.

V telekomunikacijah se kot prenosne poti v največji meri uporabljajo kovinski vodniki, optična vlakna in prazen prostor oziroma atmosfera. Preko vseh teh prenosnih poti se signali prenašajo v obliki elektromagnetnega valovanja.

Ker nas v okviru te knjige fizikalno dogajanje na sami prenosni poti ne zanima, bomo analizo reducirali na preučevanje odnosa med vzrokom in posledico, to je med vhodnim signalom $x(t)$ in izhodnim signalom $y(t)$. Prenosno pot bomo zato predstavili s preprostim sistemom z enim vhodom in enim izhodom, kot je to prikazano na sliki 3.1.



Slika 3.1 – Model prenosne poti. Prenosno pot predstavimo s sistemom, ki ima en vhod in en izhod. Izhodni signal je enak transformiranemu, zakasjenemu in popačenemu, vhodnemu signalu.

S $\mathcal{T}\{ \}$ je označena transformacija signala pri prehodu. Če poznamo to transformacijo $\mathcal{T}\{ \}$, lahko vsakemu vhodnemu signalu določimo izhodni signal:

$$y(t) = \mathcal{T}\{x(t)\} \quad (3.1)$$

zato pravimo, da transformacija $\mathcal{T}\{ \}$ v celoti opisuje tak sistem.

Velik del prenosnih poti, ki jih uporabljamo v telekomunikacijah, je linearnih in časovno nespremenljivih ali pa jih vsaj v približku lahko obravnavamo kot take. Sistem je linearen, kadar je transformacija $\mathcal{T}\{ \}$ linearna:

$$\mathcal{T}[a_1x_1(t) + a_2x_2(t)] = a_1\mathcal{T}[x_1(t)] + a_2\mathcal{T}[x_2(t)] \quad (3.2)$$

To pomeni, da je odziv sistema na vsoto dveh signalov enak vsoti ozivov na posamezna signala.

Sistem je časovno nespremenljiv, kadar velja:

$$\mathcal{T}\{x(t + t_0)\} = y(t + t_0) \quad (3.3)$$

To pomeni, da se sistem s časom ne spreminja in da je njegov odziv vedno enak.

3.1 Sistemska funkcija

Označimo s $h(t)$ odziv sistema na enotin impulz $\delta(t)$. Odziv na enotin impulz je signal na izhodu sistema, ko na njegov vhod pripeljemo enotin impulz:

$$h(t) = \mathcal{T}\{\delta(t)\} \quad (3.4)$$

Odziv sistema na enotin impulz imenujemo tudi sistemska funkcija, saj v celoti opisuje sistem. Če poznamo sistemska funkcijo, lahko namreč vsakemu vhodnemu signalu določimo izhodni signal po enačbi:

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau) d\tau = x(t) * h(t) \quad (3.5)$$

oziroma Odziv linearne, časovno nespremenljivega sistema na poljuben vhodni signal je enak konvoluciji vhodnega signala in sistemske funkcije.

3.1.1 Prevajalna funkcija

Fourierov transform sistemske funkcije imenujemo prevajalna funkcija:

$$H(f) = \mathcal{F}\{h(t)\} \quad (3.6)$$

Če naredimo Fourierov transform na obeh straneh enačbe 3.5, dobimo:

$$Y(f) = X(f) \cdot H(f) \quad (3.7)$$

Spekter signala se pri prehodu skozi linearen časovno nespremenljiv sistem preprosto množi s prevajalno funkcijo sistema.

Absolutno vrednost prevajalne funkcije imenujemo tudi amplitudna karakteristika sistema:

$$A(f) = |H(f)| \quad (3.8)$$

Argument prevajalne funkcije pa imenujemo fazna karakteristika sistema:

$$\Phi(f) = \arg\{H(f)\} \quad (3.9)$$

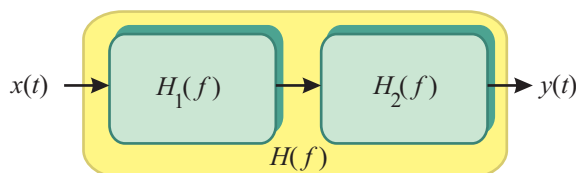
Amplitudni spekter signala se pri prehodu skozi sistem množi z amplitudno karakteristiko sistema, faznemu spektru pa se prišteje fazna karakteristika sistema.

Pri prehodu signala skozi linearen, časovno nespremenljiv sistem, se različne frekvence v signalu različno oslabijo oziroma ojačijo, zasuče pa se jim lahko tudi faza. Linearne časovno nespremenljive sisteme zato imenujemo tudi frekvenčna sita ali filtri, saj presejejo oziroma filtrirajo različne frekvence

Če dva linearna časovno nespremenljiva sistema vežemo v serijo, kot je to prikazano na sliki 3.2 je njuna skupna prevajalna funkcija enaka produktu posameznih prevajalnih funkcij:

$$H(f) = H_1(f) \cdot H_2(f) \quad (3.10)$$

Če naredimo inverzni Fourierov transform na obeh straneh enačbe (3.10), dobimo:



Slika 3.2 – Serijska vezava sistemov. Prevajalna funkcija celotnega sistema je enaka produktu posameznih prevajalnih funkcij.

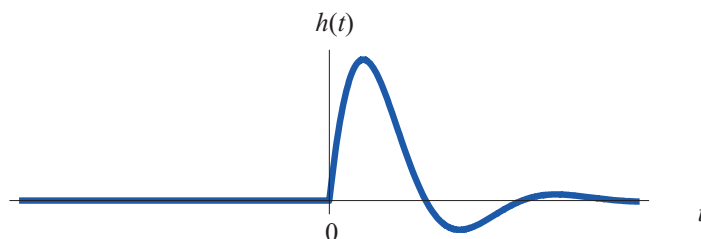
$$h(t) = h_1(t) * h_2(t) \quad (3.11)$$

Skupna sistemska funkcija je enaka konvoluciji posameznih sistemskih funkcij.

3.2 Kavzalni sistemi

Kavzalni sistemi so sistemi, pri katerih je izhod posledica vhoda. Ker posledica ne more prehitovati vzroka, pri teh sistemih izhodni signal nastopi kasneje ali kvečjemu sočasno z vhodnim signalom.

Sistemska funkcijo smo definirali kot odziv sistema na enotin impulz. Enotin impulz nastopi v času $t = 0$, zato odziv nanj ne more nastopiti pred tem časom. Sistemska funkcija kavzalnega sistema je zato za vse čase $t < 0$ enaka nič. Primer sistemske funkcije kavzalnega sistema je prikazan na sliki 3.3



Slika 3.3 – Sistemska funkcija kavzalnega sistema je enaka nič za vse čase manjše od nič, za čase večje od nič pa je lahko poljubna.

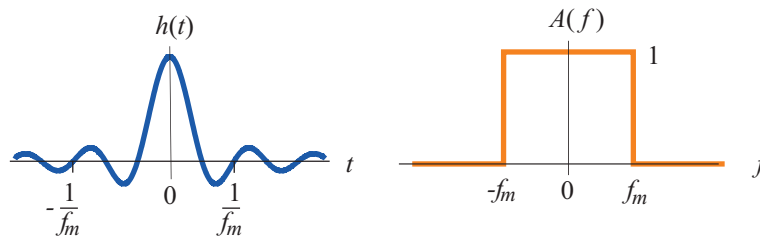
3.3 Idealno nizko sito

Idealno nizko sito imenujemo sistem ki prepušča vse frekvence do neke mejne frekvence f_m , vse frekvence nad to frekvenco pa popolnoma zaduši. Amplitudna karakteristika idealnega sita je prikazana na desni strani slike 3.4.

Če predpostavimo, da sistem ne suče faze ($\phi(f) = 0$) potem dobimo sistemsko funkcijo idealnega nizkega sita z inverznim Fourierovim transformom njegove amplitudne karakteristike:

$$h(t) = \int_{-f_m}^{f_m} e^{j2\pi ft} df = 2f_m \frac{\sin(2\pi f_m t)}{2\pi f_m t} \quad (3.12)$$

Sistemska funkcija idealnega nizkega sita je prikazana na levi strani slike 3.4. Vidimo, da je



Slika 3.4 – Sistemska funkcija in amplitudna karakteristika idealnega nizkega sita. Sistemska funkcija ni kavzalna, zato idealno nizko sito ni izvedljivo.

sistemska funkcija tudi pri časih $t < 0$ različna od nič, torej ni kavzalna. Idealno sito torej ne obstaja. Če sistemsko funkcijo premaknemo v desno (zakasnimo) ne spremenimo amplitudne karakteristike. Vendar bi morali sistemsko funkcijo premakniti neskončno v desno, da bi postal sistem kavzalen. Idealno sito bi imelo tako neskončno zakasnitev.

Z dodajanjem zakasnitve pa se idealnemu nizkemu situ lahko poljubno približamo, saj se odziv za čase manjše od nič z zakasnitvijo hitro zmanjšuje. Če ga potem postavimo na nič ne vplivamo bistveno na amplitudno karakteristiko. Več o tem je mogoče najti v literaturi o načrtovanju sit.

3.4 Močnostni spekter pri prenosu

Naključen signal $x(t)$ lahko opredelimo z njegovim močnostnim spektrom $S_x(f)$. Močnostni spekter se pri prehodu signala skozi linearen časovno nespremenljiv sistem množi s kvadratom absolutne vrednosti prevajalne funkcije sistema oziroma s kvadratom njegove amplitudne karakteristike:

$$S_y(f) = S_x(f)|H(f)|^2 = S_x(f)A^2(f) \quad (3.13)$$

V poglavju smo definirali sistemsko in prevajalno funkcijo, ki močno olajšata analizo sprememb signala pri prenosu. Ne smemo pa pozabiti, da imata sistemsko in prevajalna funkcija pomen samo pri linearnih in časovno nespremenljivih sistemih. Pri nelinearnih ali časovno spremenljivih sistemih nimata pomena. Frekvenčna sita, tudi tista izvedena z diskretnimi elementi (upori, kondenzatorji, tuljave) spadajo v skupino linearnih in časovno nespremenljivih sistemov.

Analogen prenos

Poglavje obravnava prenos signalne oblike v osnovnem pasu in višjih frekvenčnih legah. Predstavljeni so različni modulacijski postopki: amplitudna, kvadratura, enobočna, frekvenčna in fazna modulacija, ustrezni demodulacijski postopki in frekvenčno mešanje signalov.

Pri prenosu analognih signalov gre za prenos njihove oblike oziroma njihovega časovnega poteka. Idealno bi bila oblika sprejetega signala popolnoma enaka obliki oddanega signala, lahko pa je signal nekoliko zakasnen in oslabljen. Signal se pri prenosu preko neidealne prenosne poti spremeni. Govorimo o popačenju signala. Če te spremembe opazujemo v časovnem prostoru, govorimo o časovni disperziji. Če opazujemo v frekvenčnem prostoru, vidimo, da se lahko spremeni amplitudni spekter signala (amplitudno popačenje) in fazni spekter signala (fazno popačenje).

Pri prenosu signala preko linearnega sistema nastopajo samo linearna popačenja. Pri linearnem popačenju v izhodnem signalu nastopajo samo frekvence, ki so prisotne v vhodnem signalu. Če je sistem nelinearen se v izhodnem signalu pojavijo tudi nove frekvence, to je frekvence, ki niso bile prisotne v vhodnem signalu.

4.1 Prenos v osnovnem pasu

O prenosu v osnovnem pasu govorimo, kadar prenašamo signal neposredno preko prenosnega medija v nespremenjeni obliki. Tak način prenosa, čeprav vedno redkeje, se uporablja predvsem za prenos govornega signala preko kovinskega vodnika.

Spekter govornega signala pri telefonski kvaliteti zavzema frekvenčni pas med 300 Hz in 3.400 Hz. Pri teh frekvencah je amplitudna karakteristika kovinskega vodnika dokaj ravna. Problem se lahko pojavi pri daljših vodnikih, kjer postane slabljenje pri višjih frekvencah občutnejše in pride do amplitudnih popačenj. Govorni signal pa ni občutljiv na fazna popačenja, saj človeško uho ne zaznava faze signala.

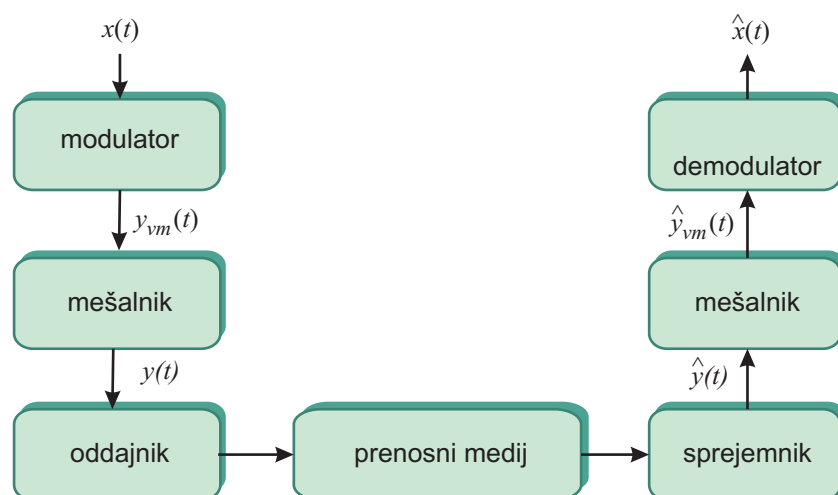
Ker se je analogni prenos govornega signala že skoraj popolnoma opustil, problematiki povezani z njim ne bomo posvečali pozornosti.

4.2 Prenos v višjih frekvenčnih legah

Večinoma analognih signalov ne prenašamo v osnovnem pasu, temveč jih preslikamo v drugo frekvenčno področje. Na ta način lahko bistveno bolje izkoriščamo prenosne poti, saj lahko na isti

prenosni poti ustvarimo več frekvenčno ločenih prenosnih kanalov po katerih lahko prenašamo med seboj neodvisne signale.

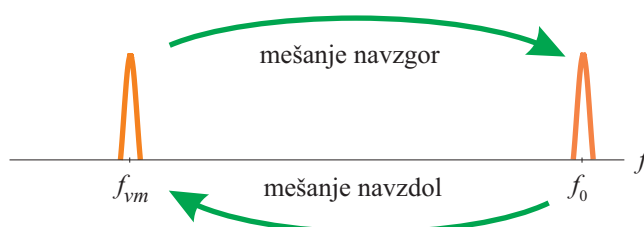
Preslikavo signala v višje frekvenčno področje dosežemo z modulatorskimi postopki in frekvenčnim mešanjem. Načelna shema prenosa v višjih frekvenčnih legah je prikazana na sliki 4.1.



Slika 4.1 – Prenos v višji frekvenčni legi. Z modulatorjem preslikamo signal na vmesno frekvenco nato pa z mešalnikom še na frekvenčni pas za prenos. Pri sprejemu sprejeti signal najprej preslikamo nazaj na vmesno frekvenco in nato z demodulatorjem nazaj v osnovni pas. Idealno bi bila na koncu signala $x(t)$ in $\hat{x}(t)$ enaka.

4.3 Frekvenčno mešanje

Z mešalnikom na levi strani slike 4.1 preslikamo signal $y_{vm}(t)$ v signal $y(t)$ tako, da se njegov spekter premakne iz enega frekvenčnega območja v drugo frekvenčno območje. Signal $y_{vm}(t)$ mora biti ozkopasoven, to pomeni, da je njegov frekvenčni pas omejen na relativno ozko frekvenčno področje okrog vmesne frekvence f_{vm} . Z mešanjem ga preslikamo v relativno ozko frekvenčno področje okrog nosilne frekvence f_0 , kot je to prikazano na sliki 4.2 To imenujemo tudi mešanje navzgor. Z mešalnikom na desni strani slike signal $\hat{y}(t)$ preslikamo iz frekvenčnega področja



Slika 4.2 – Frekvenčno mešanje navzgor. Z mešanjem prestavimo spekter signala k višjim frekvencam (mešanje navzgor) ali k nižjim frekvencam (mešanje navzdol).

okrog frekvence f_0 nazaj v signal $\hat{y}_{vm}(t)$ v frekvenčnem področju okrog frekvence f_{vm} .

4.3.1 Mešanje navzgor

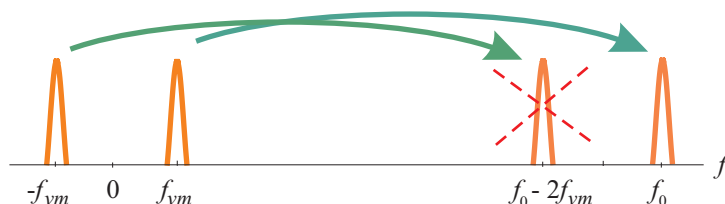
Mešanje navzgor izvedemo tako, da signal na vmesni frekvenci $y_{vm}(t)$ množimo s kosinusom razlike med nosilno frekvenco f_0 in vmesno frekvenco f_{vm} :

$$y'(t) = 2 \cdot y_{vm}(t) \cdot \cos(2\pi(f_0 - f_{vm})t) \quad (4.1)$$

V skladu z lastnostjo Fourierovega transformata podano z izrazom (2.62), je spekter signala y'_0 enak:

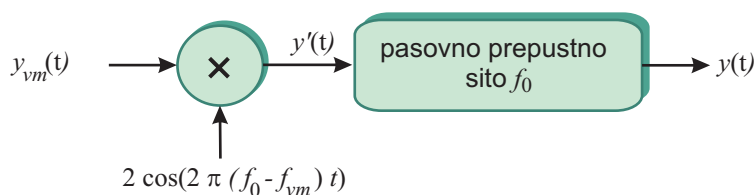
$$Y'(f) = Y_{vm}(f - f_0 + f_{vm}) + Y_{vm}(f + f_0 - f_{vm}) \quad (4.2)$$

Prvi člen vsote v zgornjem izrazu je spekter pri pozitivnih frekvencah, drugi del pa je njegova zrcalna slika pri negativnih frekvencah, kot je to prikazano na sliki 4.3.



Slika 4.3 – Mešanje navzgor. Pri množenju s kosinusom frekvence $f_0 - f_{vm}$, se spekter okrog frekvence f_{vm} preslika okrog frekvence f_0 , spekter okrog frekvence $-f_{vm}$ pa okrog frekvence $f_0 - 2f_{vm}$. Ta je odveč, zato potrebujemo še pasovno prepustno sito, ki spušča skozi samo koristen del spektra okrog frekvence f_0 .

Pri množenju se spekter pri pozitivnih frekvencah se preslika okrog frekvence f_0 , spekter pri negativnih frekvencah pa okrog frekvence $f_0 - 2f_{vm}$. Slednji je neželen del spektra, zato za množilnikom potrebujemo pasovno prepustno sito, ki prepusti zelen del spektra, neželenega pa zaduši. Šele signal na izhodu site je signal $y(t)$, ki je rezultat mešanja. Načelna vezava mešalnika navzgor je prikazana na sliki 4.4.



Slika 4.4 – Načelna vezava mešalnika navzgor. Množilniku sledi sito, ki zaduši neželene frekvenčne komponente

4.3.2 Mešanje navzdol

Pri mešanju navzdol prestavimo signal iz frekvenčnega območja okrog frekvence f_0 nazaj na območje okrog vmesne frekvenc f_{vm} . Zato signal $\hat{y}(t)$ možimo s kosinusom frekvence $f_0 - f_{vm}$.

$$\hat{y}_{vm}(t) = 2 \cdot \hat{y}(t) \cos(2\pi(f_0 - f_{vm})t) \quad (4.3)$$

Za spekter signalov tedaj velja:

$$\hat{Y}'_{vm}(f) = \hat{Y}(f + f_0 - f_{vm}) + \hat{Y}(f - f_0 + f_{vm}) \quad (4.4)$$

dročje uporabimo mešanje.

Z modulacijski postopkom moduliramo (spreminjamo nek parameter) harmonskega nosilca. Običajno privzamemo, da je to kosinusni signal v obliki:

$$A \cos(2\pi \int f dt + \phi) \quad (4.5)$$

V zgornji enačbi nastopajo trije parametri, ki jih lahko moduliramo, to so amplituda A , trenutna frekvenca f in fazni zamik ϕ . Kadar spreminjamo amplitudo, govorimo o amplitudni modulaciji, kadar pa spreminjamo frekvenco ali fazni zamik, pa govorimo o kotni modulaciji, to je frekvenčni oziroma fazni modulaciji. Ustrezen parameter spreminjamo z modulacijskim signalom $x(t)$, to je s signalom, ki ga želimo prenašati. V nadaljevanju bomo predpostavili, da je srednja vrednost modulacijskega signala enaka nič:

$$\overline{x(t)} = 0 \quad (4.6)$$

in da je normiran po amplitudi tako, da velja:

$$\max\{|x(t)|\} = 1 \quad (4.7)$$

4.5 Amplitudna modulacija

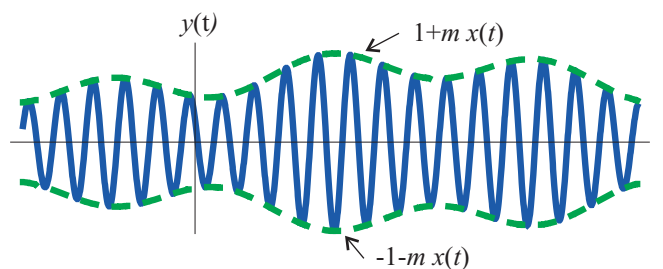
Pri amplitudni modulaciji (AM) sta frekvenca in fazni zamik konstantna. Privzamemo $f = f_{vm}$ in $\phi = 0$, z modulacijskim signalom pa spreminjamo amplitudo signala:

$$A = A(t) = 2(1 + m x(t)) \quad (4.8)$$

Ker je amplituda signala vedno pozitivna in želimo, da je po obliki enaka modulacijskemu signalu, smo modulacijski signal množili z modulacijskim indeksom m in mu prišteli 1. Modulacijski indeks je število med 0 in 1, oziroma med 0 in 100%. Pri 100% modulaciji, se amplituda moduliranega signala $y(t)$ spreminja med 2 in 0, če pa je modulacijski indeks enak 0, signala ne moduliramo, saj ostaja amplituda nosilca konstantna in enaka 1. Moduliran signal lahko sedaj zapišemo v obliki:

$$y(t) = 2(1 + m x(t)) \cos(2\pi f_{vm} t) \quad (4.9)$$

Primer amplitudno moduliranega signala je prikazan na sliki 4.7.

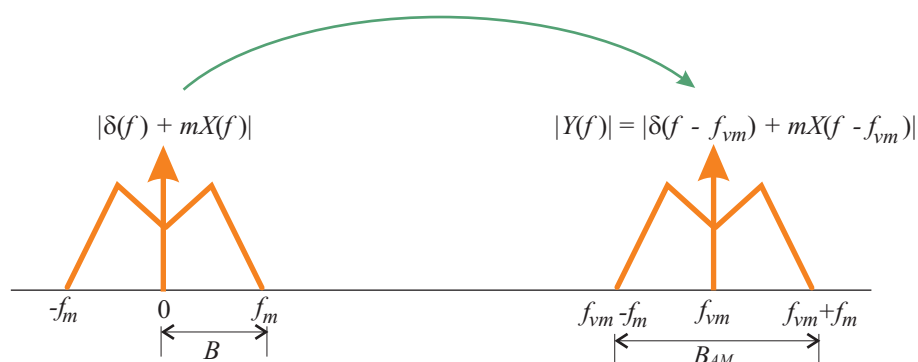


Slika 4.7 – Amplitudno moduliran signal. Modulacijski indeks je manjši od 1, saj amplituda ni nikoli enaka nič.

V skladu z lastnostjo Fourierovega transformata (2.62) preslika amplitudna modulacija spekter signala iz osnovnega pasu v pas okrog frekvence nosilca f_{vm} . Ker se poleg pozitivnih frekvenc preslikajo tudi negativne frekvence, se pasovna širina signala B med frekvenco 0 in mejno frekvenco f_m v osnovnem pasu pri modulaciji podvoji.

$$B_{AM} = 2B \quad (4.10)$$

Shematsko je to prikazano na sliki 4.8.



Slika 4.8 – Spekter amplitudne modulacije. Spekter modulacijskega signala $X(f)$ se preslika okrog frekvence nosilca. Dobimo levi in desni bočni pas moduliranega signala, zato je pasovna širina signala pri modulaciji podvojena. Enosmerna komponenta, ki smo jo prišteli modulacijskemu signalu se preslika v nosilec.

Amplitudna modulacija ima predvsem dve slabosti:

- Pasovna širina signala se pri modulaciji podvoji, zato rabimo za prenos dvakrat tolikšen frekvenčni pas kot pri prenosu v osnovnem pasu.
- Veliko moči pri prenosu se porabi za prenos nosilca, ki je prisoten v moduliranem signalu. Ker je nosilec periodičen signal ne prispeva k prenosu signalne oblike, ki je naključna. Trdimo lahko, da gre ta moč v izgubo.

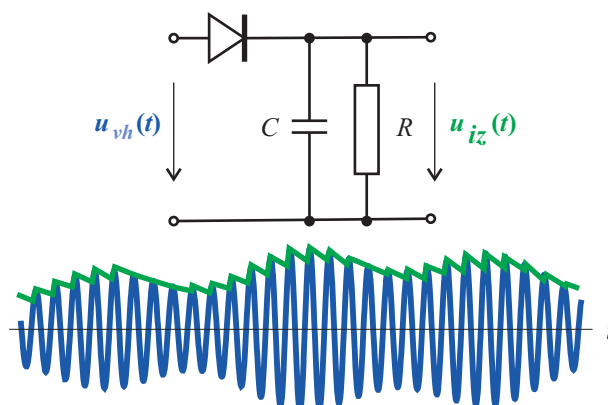
Amplitudno moduliran signal lahko demoduliramo z detektorjem ovojnice. Preprost detektor ovojnice je prikazan na sliki 4.9.

Nazobčanost signala na izhodu detektorja ovojnice povzroča v spektru detektiranega signala visoke frekvenčne komponente, ki jih preprosto izločimo z nizkim sitom. Če gre za govorni signal, ki ga poslušamo preko zvočnika, teh na zvočniku ni slišati.

4.6 Modulacija z zadušenim nosilcem

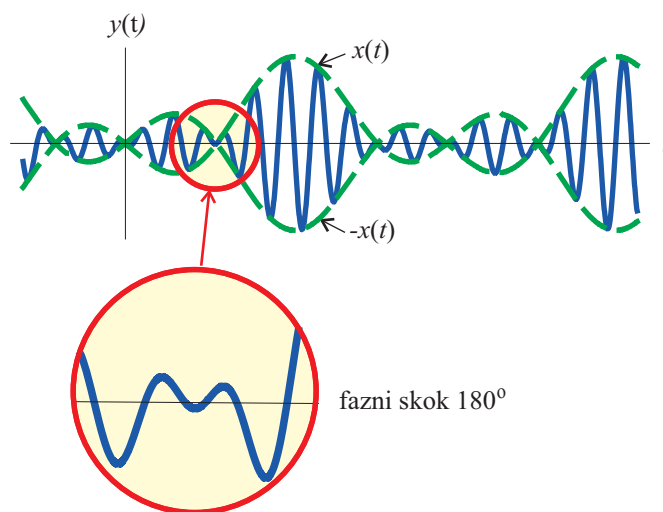
Pri modulaciji z zadušenim nosilcem (SCAM – Suppressed Carrier Amplitude Modulation) modulacijskemu signalu ne dodajamo enosmerne komponente. Nosilni signal množimo neposredno z modulacijskim signalom:

$$y(t) = 2x(t) \cos(2\pi f_{vm}t) \quad (4.11)$$



Slika 4.9 – Detektor ovojnice. Kondenzator C se polni preko diode na maksimalno vrednost signala, potem pa se prazni preko upornosti R .

S tem se izognemo izgubi moči zaradi nosilca v moduliranem signalu, vendar nam to onemogoča demodulacijo s pomočjo detekcije ovojnice. Amplituda je namreč enaka absolutni vrednosti modulacijskega signala, kot je to razvidno na sliki 4.10.



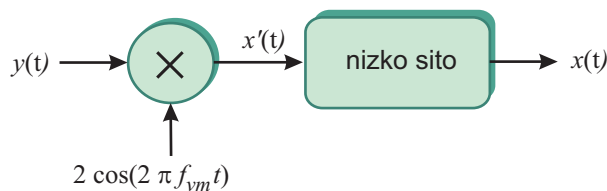
Slika 4.10 – Moduliran signal z zadušenim nosilcem. Z detekcijo ovojnice bi dobili absolutno vrednost modulacijskega signala. Vendar nastopi v moduliranem signalu ob prehodu modulacijskega signala skozi nič fazni skok 180° , ki ga izkoristimo pri sinhroni demodulaciji.

Spekter signala se pri SCAM preslika popolnoma enako kot pri AM (slika 4.8). Ker pa v modulatorskem signalu ni enosmerne komponente v moduliranem signalu nosilec ni več prisoten. Pasovna širina se še vedno podvoji:

$$B_{SCAM} = 2B \quad (4.12)$$

SCAM moduliran signal lahko demoduliramo s sinhronim demodulatorjem, kot je to prikazano na sliki 4.11

Pri sinhroni demodulaciji potrebujemo nosilni signal, ki natanko ustreza signalu, ki smo ga modulirali, tako po amplitudi kot po fazi. Da bi ga lahko pri sprejemu dobili, lahko v moduliran



Slika 4.11 – Sinhroni demodulator. Moduliran signal ponovno množimo z nosilcem. Pri tem se spekter preslika okrog frekvence nič in okrog dvojne frekvence. Dvojno frekvenco izločimo z nizkim sitom.

signal dodamo majhen nosilec, ki ga pri sprejemu izločimo s sitom, ali pa lahko levo in desno od dodamo dva pilotna nosilca iz katerih je z mešanjem mogoče dobiti nosilec. V obeh primerih pa je moč, ki jo za to potrebujemo manjša kot pri amplitudni modulaciji.

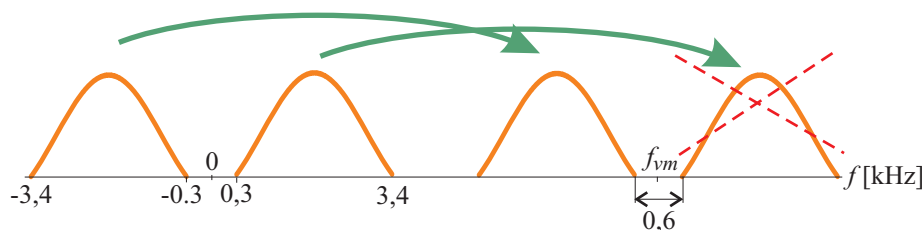
4.7 Enobočna modulacija

Ker sta levi in desni bočni pas pri AM in SCAM le zrcalni sliki (levi so preslikane negativne, desni pa preslikane pozitivne frekvence), zadošča, da prenašamo samo en bočni pas. Tako modulacijo zato imenujemo enobočna modulacija (SSB - Single Side Band). Pri enobočni modulaciji lahko prenašamo levi ali desni bok. Postopka modulacije in demodulacije sta popolnoma enaka pri SCAM, s to razliko, da pri SSB po množenju zadušimo en bok s ustreznim sitom.

$$y'(t) = 2x(t) \cos(2\pi f_{vm}t) \quad (4.13)$$

SSB modulacija je pravzaprav ekvivalentna mešanju. Edina razlika je v tem, da je pri mešanju vhodni signal ozkopasoven signal na vmesni frekvenci. Zato sta pri mešanju levi in desni bok razmaknjena in lahko en bok zadušimo z dokaj nezahtevnim sitom.

Enobočna modulacija se uporablja predvsem pri prenosu govornega signala, ki zavzema frekvenčno področje med 300 Hz in 3.400 Hz. Pri množenju z nosilcem sta levi in desni bok tako razmaknjena za 600 Hz, kot je to prikazano na sliki 4.12.



Slika 4.12 – Enobočna modulacija govornega signala. Signal moramo modulirati na dokaj nizko vmesno frekvenco, da lahko potem z ustreznim sitom izločimo samo en bok.

Vmesno frekvenco moramo izbrati dovolj nizko, da lahko izločimo en bočni pas, saj je na nizkih frekvencah frekvenčni razmik med bočnima pasovama relativno precej večji kot na visokih frekvencah. Za prenos pri zelo visokih frekvencah je zato lahko potrebno dvakratno mešanje.

Pri SSB se pasovna širina signala ne poveča:

$$B_{SSB} = B \quad (4.14)$$

zato rabimo za prenos polovico manjšo pasovno širino kot pri AM oziroma pri SCAM.

4.8 Fazna modulacija

Pri fazni modulaciji (PM – Phase Modulation) sta amplituda in frekvenca konstantni ($A = 1$, $f = f_{vm}$), z modulatorskim signalom pa spreminjamo fazni zamik:

$$\phi = \phi(t) = \Delta\phi x(t) = 2\pi m x(t) \quad (4.15)$$

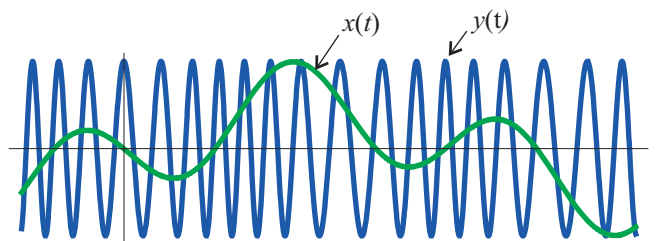
kjer je m modulatorski indeks fazne modulacije:

$$m = \frac{\Delta\phi}{2\pi} \quad (4.16)$$

ker je v zgornji enačbi signal $x(t)$ normiran, spreminjamo fazo za $\pm\Delta\phi$ oziroma za m period nosilca. Fazno moduliran signal lahko sedaj zapišemo v obliki:

$$y(t) = \cos(2\pi(f_{vm}t + m x(t))) \quad (4.17)$$

Fazno moduliran signal je prikazan na sliki 4.13



Slika 4.13 – Fazno moduliran signal. Faza signala se spreminja s časom. Zaradi spreminjanja faze se spreminja tudi trenutna frekvenca. Kadar je naklon modulatorskega signala pozitiven je frekvenca visoka, kadar pa je negativen, je frekvenca nizka.

Fazna modulacija je nelinearen postopek, zato spektra fazno moduliranega v splošnem ni mogoče določiti analitično. Odvisen je tako od pasovne širine signala in modulatorskega indeksa. Pasovno širino fazno moduliranega signala lahko približno ocenimo po enačbi:

$$B_{PM} = 2B(m + 1) \quad (4.18)$$

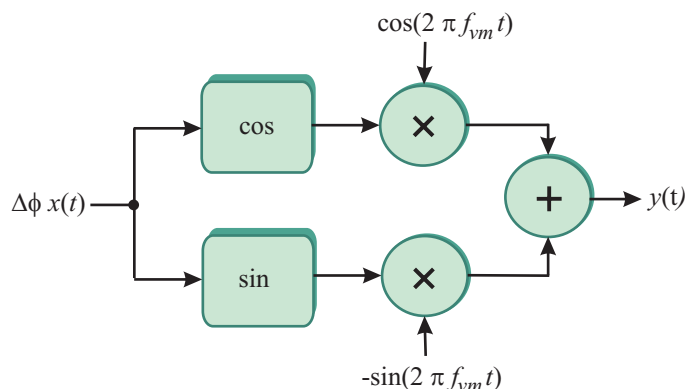
Spekter se z večanjem modulatorskega indeksa širi. Pri fazni modulaciji je modulatorski indeks običajno relativno majhen ($m < 1/2$), zato spekter fazno moduliranega signala ni bistveno širši od spektra amplitudno moduliranega signala.

Pri izvedbi faznega modulatorja lahko upoštevamo enakost:

$$\cos(2\pi f_{vm}t + \Delta\phi x(t)) = \cos(\Delta\phi x(t)) \cos(2\pi f_{vm}t) - \sin(\Delta\phi x(t)) \sin(2\pi f_{vm}t) \quad (4.19)$$

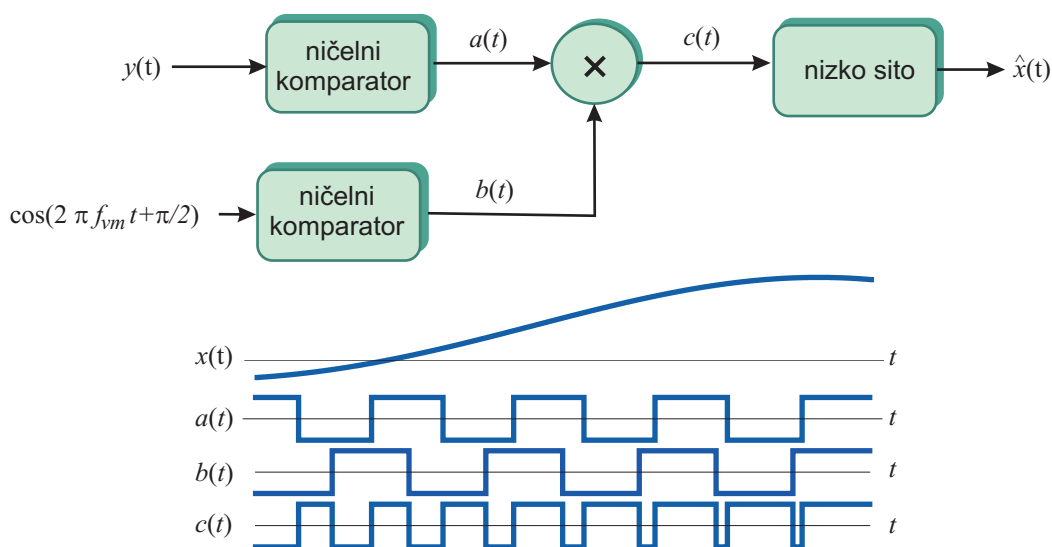
kot je to prikazano na sliki 4.14.

Demodulacijo fazno moduliranega signala lahko izvedemo s faznim detektorjem. Fazni detektor primerja modulirani signal in za 90° fazno zameknjen nosilec. Oba signala najprej vodimo



Slika 4.14 – Fazni modulator izveden s pomočjo tako imenovanega kvadraturnega modulatorja. Kosinus in sinus faze množimo z nosilcema, ki sta v kvadraturi, to je ortogonalna.

preko ničelnega komparatorja in ju nato množimo. Če sta nosilec in moduliran signal v fazi, potem sta pravokotna signala na vходу množilnika zamaknejna za 90^0 . Srednja vrednost produkta je tedaj enaka nič. Ko modulacijskemu signalu spreminjamo fazo, se spreminja tudi srednja vrednost signala na izhodu množilnika. Z nizkim sitom izločimo visokofrekvenčne komponente, tako da ostane le počasi se spreminjajoča srednja vrednost, ki je proporcionalna faznemu zamiku moduliranega signala. Fazni modulator, modulacijski signal in signali na vhidih in izhodu množilnika so prikazani na slik 4.15



Slika 4.15 – Fazni detektor. Za 90^0 fazno zamaknjen nosilec in fazno moduliran signal vodimo preko komparatorja na množilnik. Nizko sito izloči visokofrekvenčne komponente signala na izhodu množilnika, tako da ostane samo nizkofrekvenčna komponenta, ki je sorazmerna fazni razliki obeh signalov.

S faznim detektorjem lahko detektiramo samo fazne razlike med -90^0 in 90^0 . Pri demodulaciji s faznim detektorjem zato $\Delta\phi$ ne presega 90^0 , oziroma mora biti modulacijski indeks manjši od ene četrtine ($m < 1/4$).

Fazno moduliran signal lahko demoduliramo tudi s kvadraturnim demodulatorjem, o čemer

bo nekoliko več govora pri digitalnih modulacijskih postopkih v naslednjem poglavju.

4.9 Frekvenčna modulacija

Pri frekvenčni modulaciji (FM) sta amplituda in fazni zamik konstantna ($A = 1$, $\phi = 0$), z modulacijskim signalom pa spreminjamo trenutno frekvenco signala:

$$f(t) = f_{vm} + \Delta f x(t) = f_{vm} + mBx(t) \quad (4.20)$$

kjer je m modulacijski indeks frekvenčne modulacije:

$$m = \frac{\Delta f}{B} \quad (4.21)$$

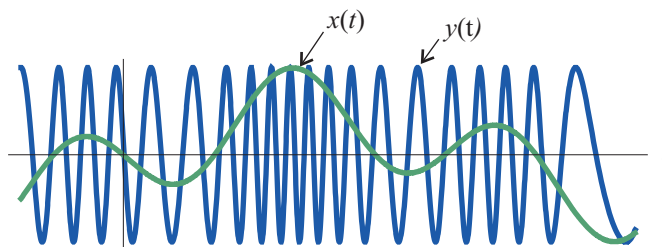
Frekvenčno moduliran signal zapišemo v obliki:

$$y(t) = \cos \left(2\pi \left(f_{vm}t + mB \int x(t) dt \right) \right) \quad (4.22)$$

Kljub temu, da pri fazni modulaciji spreminjamo fazni zamik, pri frekvenčni modulaciji pa trenutno frekvenco signala, lahko opazimo, da sta si enačbi za fazno moduliran signal (4.17) in frekvenčno moduliran signal (4.22) zelo podobni. Razlika je predvsem v tem, da v drugi nastopa modulacijski signal $x(t)$ pod integralom v prvi pa ne.

Frekvenčna in fazna modulacija se v bistvu ne razlikujeta. Frekvenčno moduliran signal lahko dobimo, če na fazni modulator pripeljemo integral modulacijskega signala. Podobno lahko dobimo fazno moduliran signal tako, da na vhod frekvenčnega modulatorja pripeljemo odvod modulacijskega signala.

Primer frekvenčno moduliranega signala z istim modulacijskim signalom kot pri fazni modulaciji, je prikazana sliki 4.16



Slika 4.16 – Frekvenčno moduliran signal. Pri visokih amplitudah modulacijskega signala je frekvenca moduliranega signala visoka, pri nizkih pa nizka.

Podobno kot fazna modulacija je tudi frekvenčna modulacija nelinearen proces, zato tudi ne moremo analitično določiti spektra moduliranega signala, na enak način pa lahko približno ocenimo pasovno širino moduliranega signala:

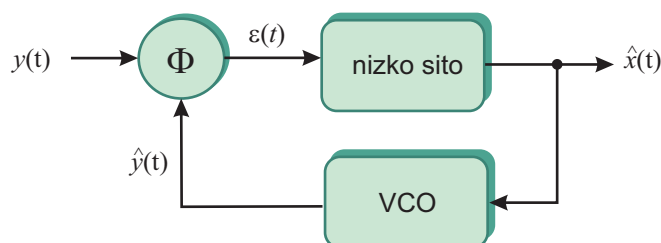
$$B_{FM} = 2(\Delta f + B) = 2B(m + 1) \quad (4.23)$$

Za razliko od fazne modulacije pa je pri frekvenčni modulaciji modulaijski indeks lahko zelo velik ($m \gg 1$), zato je lahko tudi spekter bistveno širši kot pri amplitudni modulaciji.

Pri velikem modulaijskem indeksu se trenutna frekvenca spreminja relativno počasi, zato lahko ocenimo tudi obliko spektra frekvenčno moduliranega signala z velikim modulaijskim indeksom. Pri frekvenčni modulaciji namreč določena amplituda (trenutna vrednost) modulaijskega signala ustreza določeni frekvenci v moduliranem signalu. V spektru moduliranega signala bodo imele torej višjo amplitudo frekvenčne komponente, ki ustrezajo pogostejšim amplitudam modulaijskega signala. Amplitudni spekter frekvenčno moduliranega signala bo zato po obliki približno enak amplitudni porazdelitvi modulaijskega signala. Amplitudna porazdelitev signala namreč pove, kako pogosto nastopajo posamezne amplitude v signalu.

4.9.1 Demodulacija FM s fazno ujeto zanko

Demodulacijo frekvenčno moduliranega signala lahko izvedemo s fazno ujeto zanko (PLL – Phase Locked Loop). Fazno ujeto zanko izvedemo z napetostno krmiljenim oscilatorjem (VCO) v povratni vezavi, kot je to prikazano na sliki 4.17.



Slika 4.17 – Fazno ujeta zanka je VCO vezan v povratno vezavo. S faznim detektorjem Φ primerjamo fazo frekvenčno moduliranega signala in signala $\hat{y}(t)$ na izhodu VCO. S signalom napake, to je razlika med fazama signalov, $\epsilon(t)$ preko nizkega sita krmilimo VCO tako, da faza $\hat{y}(t)$ sledi fazi vhodnega signala $y(t)$. Pravimo da sta signala fazno zaklenjena. Ko sta signala fazno zaklenjena imata tudi isto trenutno frekvenco. Ker je trenutna frekvenca $\hat{y}(t)$ enaka trenutni frekvenci $y(t)$, mora biti tudi signal $\hat{x}(t)$ s katerim krmilimo VCO enak signalu $x(t)$, s katerim smo krmilili VCO pri modulaciji.

Fazni detektor primerja fazi vhodnega frekvenčno moduliranega signala $y(t)$ in signala na izhodu VCO $\hat{y}(t)$. Razlika v fazi tvori signal napake $\epsilon(t)$, s katerim krmilimo VCO tako, da zmanjšujemo signal napake. Ko sta fazi obeh signalov enaki, pravimo, da sta signala fazno zaklenjena. Fazno zaklenjena signala imata enako trenutno frekvenco. Ker ima signal $\hat{y}(t)$ na izhodu VCO enako trenutno frekvenco, kot frekvenčno moduliran signal $y(t)$, mora biti signal na vhodu VCO $\hat{x}(t)$ enak signalu $x(t)$ s katerim smo krmilili VCO pri modulaciji.

Pri fazno ujeti zanki sta pomembna predvsem dva parametra, to sta območje lovljenja in območje zaklepanja.

Območje lovljenja je območje frekvenc vhodnega signala v katerem se fazno ujeta zanka lahko ujame. To območje je odvisno od nizkega sita, ki je vključeno v zanko. Frekvenca signala napake je namreč enaka razliki frekvenc vhodnega signala in signala na izhodu VCO. Če je ta frekvenca nad mejno frekvenco nizkega sita (če je razlika frekvenc prevelika), sito zaduši signal napake s

katerim krmilimo VCO, zato ne pride do faznega zaklepanja. Če želimo široko območje lovljenja mora biti torej mejna frekvenca nizkega sita visoka.

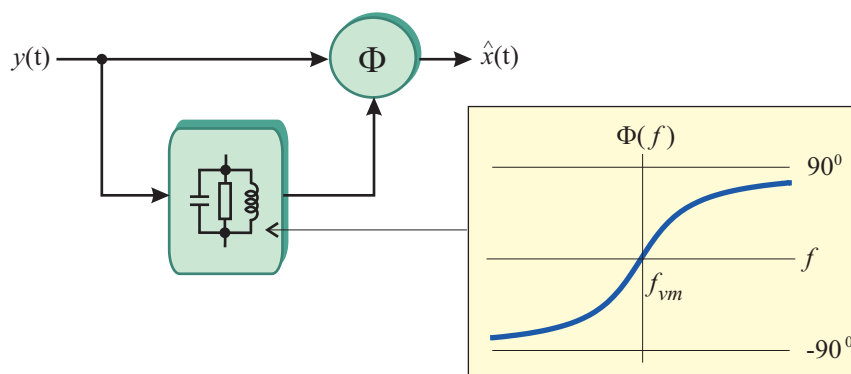
Po drugi strani pa visoka mejna frekvenca sita omogoča hitre spremembe krmilnega signala VCO, kar povzroča hitre spremembe faze ali tako imenovano tresenje faze. Načrtovanje nizkega sita v fazno ujete zanki zato predstavlja kompromis med širino območja lovljenja in tresenjem faze.

Območje zaklepanja je območje frekvenc vhodnega signala, v katerem lahko fazno ujeta zanka obdrži signala zaklenjena ko sta že zaklenjena. To območje je odvisno predvsem od frekvenčnega območja VCO.

Fazno ujeta zanka je lahko zaradi povratne vezave nestabilna. Nestabilna je kadar je povratna povezava pozitivna (razlika v fazi povzroči še večjo razliko v fazi) in ne negativna (razlika v fazi povzroči zmanjšanje razlike), kot bi morala biti. Tudi na stabilnost fazno ujete zanke dosežemo z ustreznim načrtovanjem nizkega sita.

4.9.2 Demodulacija s FM s sukanjem faze in faznim detektorjem

Za demodulacijo FM signala najpogosteje uporabljamo demodulator, ki temelji na frekvenčno odvisnem sukanju faze moduliranega signala in faznem detektorju, kot je to prikazano na sliki 4.18.



Slika 4.18 – Demodulator sukanjem faze in faznim detektorjem. Fazo vhodnega signala najprej zasučemo z nihajnim krogom, ki ima v območju okrog centralne frekvence dokaj linearno fazno karakteristiko. S faznim detektorjem nato primerjamo fazo originalnega in fazno zasukanega signala. Ker je fazni zasuk odvisen od trenutne frekvence frekvenčno moduliranega signala, je signal na izhodu enak modulacijskemu signalu.

Moduliranemu signalu najprej zasučemo fazo sorazmerno s frekvenco in nato primerjamo fazi moduliranega signala in signala z zasukano fazo. Razlika v fazah je enaka trenutni frekvenci signala. Za sukanje faze lahko uporabimo kar nihajni krog, ki ima okrog centralne frekvence dokaj linearno karakteristiko.

Frekvenčna modulacija je bistveno manj občutljiva na šum kot amplitudna modulacija, zato se frekvenčna modulacija uporablja, kadar želimo imeti kvaliteten prenos, npr. pri prenosu radijskih programov. FM radijski programi oddajajo na frekvenčnem področju med 90 Mhz in 110 Mhz, vmesna frekvenca je 10,7 Mhz, posamezni kanali zavzemajo pasovno širino 75 kHz in

so med seboj razmaknjeni za 100 kHz.

Čeprav imata frekvenčna in fazna modulacija širši spekter kot amplitudna modulacija, pa je njuna prednost poleg večje odpornosti proti šumu tudi v tem, da amplituda signala ne nosi informacije. Zato lahko na oddajniku uporabimo nelinearne ojačevalnike, ki imajo bistveno boljši izkoristek kot linearni.

Digitalni signali

V poglavju so na kratko predstavljeni digitalni signali oziroma časovni nizi števil. Predstavljeno je vzorčenje signalov in podan je teorem o vzorčenju. Predstavljeni so postopki za spektralno analizo digitalnih signalov.

Digitalen (številski) signal je niz števil, ki si sledijo v časovnem zaporedju. Digitalne signale označujemo kot funkcije indeksa, kot na primer $x[n]$, $y[n]$. Indeks n je celo število, ki pove mesto določenega števila v nizu. Ker gre za časovno zaporedje, predstavlja indeks n časovni indeks. Oglate oklepaje uporabljamo, da ločimo digitalne signale od analognih signalov, kjer smo uporabljali okrogle oklepaje.

Digitalen signal lahko predstavlja določene podatke, kot so to na primer tekstovne datoteke, lahko pa so to vzorci analognega signala.

5.1 Vzorčenje signalov

Digitalen signal $x[n]$ lahko dobimo z vzorčenjem analognega signala $x(t)$. Najpogosteje vzorčimo v enakomernih časovnih intervalih T_{vz} oziroma s konstantno vzorčno frekvenco f_{vz} , kjer velja:

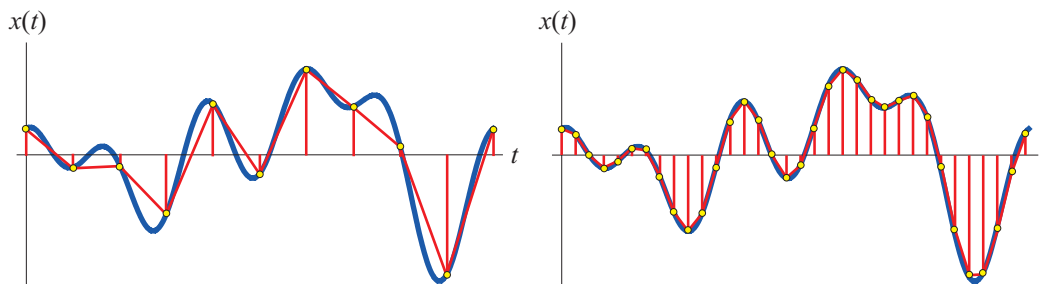
$$f_{vz} = \frac{1}{T_{vz}} \quad (5.1)$$

Pri enakomernem vzorčenju je digitalni signal $x[n]$ enak vrednostim analognega signala $x(t)$ ob vzorčnih trenutkih:

$$x[n] = x(nT_{vz}) \quad (5.2)$$

Vzorčimo lahko različne analogne signale, kot sta to na primer govorni in avdio signal. Vzorčimo jih zato, da jih lahko shranjujemo ali prenašamo digitalno. Na zgoščenkah je avdio signal zapisan v digitalni obliki. V sodobni telefoniji govorni signal prenašamo digitalno. V obeh primerih želimo na koncu iz vzorcev signala rekonstruirati originalni signal, da bi ga lahko predvajali preko zvočnika ali slušalk.

V splošnem iz vzorcev signala ne moremo enoumno rekonstruirati originalnega signala, saj vedno obstaja neskončna množica signalov, ki ima ob vzorčnih trenutkih iste vrednosti. Originalni signal lahko zelo dobro rekonstruiramo iz njegovih vzorcev, če je le frekvenca vzorčenja dovolj visoka glede na hitrost s katero se spreminja signal. Pri visoki vzorčni frekvenci, lahko dobimo dober približek originalnega signala že s preprosto linearno interpolacijo (povezavo po dveh sosednjih vzorcih), kot je to prikazano na sliki 5.1.



Slika 5.1 – Pri prenizki vzorčni frekvenci se signal, ki ga dobimo z linearno interpolacijo močno razlikuje od originalnega, če pa je vzorčna frekvenca dovolj visoka, dobimo z linearno interpolacijo dober približek originalnega signala.

5.1.1 Teorem o vzorčenju

Linearna interpolacija ni najboljši način za rekonstrukcijo signala. Pri linearni interpolaciji mora biti namreč vzorčna frekvenca relativno visoka, da bi dobili dober približek originalnega signala. To pomeni, da bi za shranjevanje signala v digitalni obliki potrebovali veliko pomnilniškega prostora oziroma veliko prenosno hitrost za njegov prenos. Poleg tega da linearna interpolacija samo približek signala in ne omogoča popolne rekonstrukcije.

Veliko boljši rezultat lahko dobimo s tako imenovano nizkopasovno interpolacijo. Videli smo, da je potrebna hitrost vzorčenja odvisna od hitrosti spreminjanja signala. Hitrost spreminjanja signala lahko omejimo tako, da omejimo njegov spekter. Signal se namreč ne more spreminjati hitreje kot se spreminja njegova najvišja frekvenčna komponenta.

Naj bo spekter $X(f)$ signala $x(t)$ frekvenčno omejen z mejno frekvenco f_m , kar pomeni da ne vsebuje frekvenčnih komponent nad to frekvenco. Vzorčenje predstavimo kot množenje z vzorčevalno funkcijo $v_\delta(t)$, ki je niz enotinih impulzov, razmaknjenih za T_{vz} :

$$v_\delta(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_{vz}) \quad (5.3)$$

Z enotinimi impulzi vzorčen signal $x_\delta(t)$ je potem enak:

$$x_\delta(t) = x(t) \cdot v_\delta(t) \quad (5.4)$$

kjer so z $x[n]$ označeni vzorci signala $x(t)$.

Vzorčevalna funkcija je periodičen signal z osnovno frekvenco f_{vz} . V skladu z lastnostjo Fourierovega transformata (2.60), je njen spekter enak:

$$V_\delta(f) = \frac{1}{T_{vz}} \sum_{k=-\infty}^{\infty} \delta(f - kf_{vz}) \quad (5.5)$$

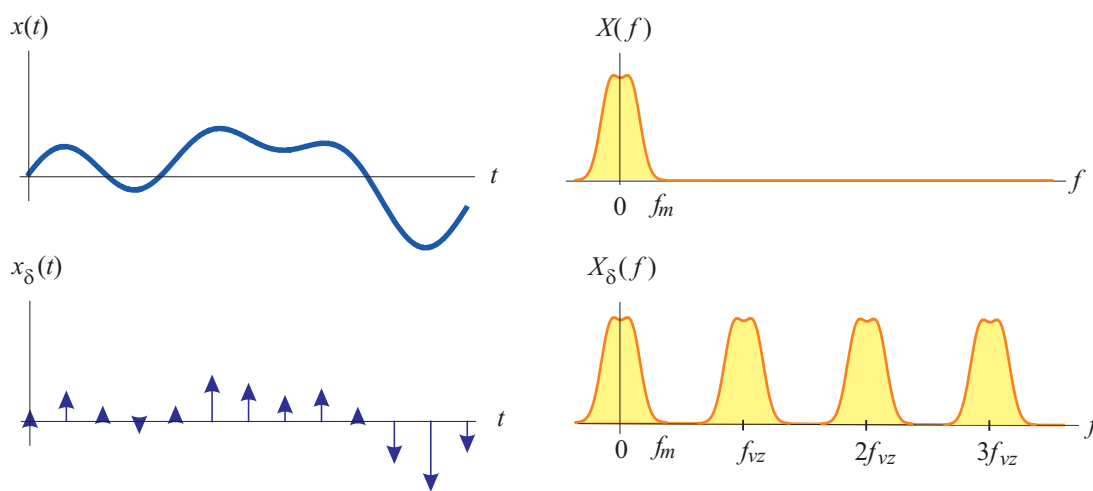
Naredimo Fourierov transform na obeh straneh enačbe (5.4) in upoštevamo, da je Fourierov transform produkta enak konvoluciji:

$$X_\delta(f) = X(f) * V_\delta(f) = \frac{1}{T_{vz}} \sum_{k=-\infty}^{\infty} X(f) * \delta(f - kf_{vz}) \quad (5.6)$$

Ker je konvolucija spektra signala s frekvenčno prmaknjenim enotnim impulzom enaka premaknjenemu spektru signala, lahko zapišemo:

$$X_{\delta}(f) = \frac{1}{T_{vz}} \sum_{k=-\infty}^{\infty} X(f + kf_{vz}) \quad (5.7)$$

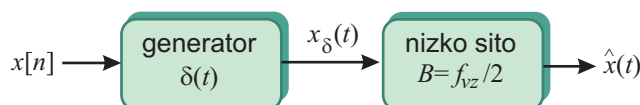
Pri vzorčenju z enotnimi impulzi se spekter originalnega signala preslika okrog mnogokratnikov vzorčne frekvence kf_{vz} , kot je to prikazana na sliki 5.2 Vprašanje, ali možno iz vzorcev signala



Slika 5.2 – Pri vzorčenju z enotnimi impulzi se spekter originalnega signala preslika okrog mnogokratnikov vzorčne frekvence. Da se preslikani spektri med seboj ne prekrivajo, mora biti vzorčna frekvenca vsaj dvakrat višja od mejne frekvence signala.

rekonstruirati originalni signal, je ekvivalentno vprašanju, ali je možno iz spektra vzorčenega signala $X_{\delta}(f)$ dobiti nazaj spekter originalnega signala $X(f)$.

Če je frekvenca vzorčenja f_{vz} vsaj dvakrat višja od mejne frekvence signala f_m , potem se ponovljeni spektri med seboj ne prekrivajo in je prva ponovitev spektra okrog frekvence nič enaka spektru originalnega signala. Da bi iz delta vzorčenega signala dobili nazaj originalni signal moramo iz spektra vzorčenega signala odstraniti vse ponovitve spektra, kar lahko naredimo z idealnim nizkim sitom s pasovno širino $B = f_{vz}/2$. Originalen signal lahko torej rekonstruiramo tako, da najprej iz vzorcev $x[n]$ tvorimo vzorčeni signal $x_{\delta}(t)$ in ga nato vodimo na nizko sito, kot je to prikazano na sliki 5.3



Slika 5.3 – Rekonstrukcija signala iz vzorcev. Z generatorjem enotnih impulzov najprej tvorimo vzorčeni signal, ki ga vodimo na nizko sito z meno frekvenco $f_{vz}/2$. Če pri vzorčenju ni prišlo do prekrivanja spektrov ($f_{vz} \geq 2f_m$) potem je rekonstruiran signal $\hat{x}(t)$ na izhodu sita enak originalnemu signalu $x(t)$.

Teorem o vzorčenju. Signal je mogoče popolnoma rekonstruirati iz njegovih vzorcev, če je vzorčna frekvenca vsaj dvakrat višja od najvišje frekvence v signalu. Rekonstruiramo ga tako,

da iz z enotinimi impulzi vzorčenega signala z nizkim sitom izločimo vse okrog mnogokratnikov vzorčne frekvence preslikane spektre.

V praksi mora biti vzorčna frekvenca višja od dvojne mejne frekvence, ker bi drugače potrebovali idealno sito, ki pa ni izvedljivo. Govorni signal z mejno frekvenco 3,4 kHz na primer vzorčimo z vzorčno frekvenco 8 kHz, audio signal z mejno frekvenco 20 kHz pa vzorčimo s vzorčno frekvenco 44,1 kHz.

5.2 Spekter digitalnih signalov

V prejšnjem razdelku smo spekter z enotinimi impulzi vzorčenega signala izrazili s spektrom originalnega signala. Lahko pa ga izrazimo neposredno z vzorci signala. Vzorec signala $x_\delta(t)$ lahko zapišemo v obliki:

$$x_\delta(t) = \sum_{n=-\infty}^{\infty} x[n]\delta(t - nT_{vz}) \quad (5.8)$$

Če naredimo Fourierov transform na obeh straneh zgornjega izraza, dobimo:

$$X_\delta(f) = \sum_{n=-\infty}^{\infty} x[n]e^{-2\pi n f T_{vz}} \quad (5.9)$$

Spekter digitalnega signala $x[n]$ definiramo tako, da je enak spektru ustreznega z enotinimi impulzi vzorčenega signala, le da namesto frekvence f pri diskretnih signalih uporabljamo normirano frekvenco F :

$$F = \frac{f}{f_{vz}} = f \cdot T_{vz} \quad (5.10)$$

Spekter $X(F)$ časovno diskretnega signala $x[n]$ je potem enak:

$$X(F) = X_\delta(F \cdot f_{vz}) = \sum_{n=-\infty}^{\infty} x[n]e^{-2\pi n F} \quad (5.11)$$

Zgornji izraz imenujemo časovno diskretni Fourierov transform (TDFT – Time Discrete Fourier Transform). Spekter digitalnega signala je periodičen po frekvenci F . Kadar so to vzorci analognega signala predstavlja frekvenca $F = 1$ vzorčno frekvenco. Ker morajo biti signali pred vzorčenjem frekvenčno omejeni na $f_m \leq F_{vz}/2$, je spekter zanimiv le v frekvenčnem področju $0 \leq F \leq 0,5$.

Močnostni spekter časovno diskretnega signala je definiran kot TDFT avtokorelacijske funkcije $R_{xx}[m]$:

$$R_{xx}[m] = \lim_{N \rightarrow \infty} \left(\frac{1}{2N} \sum_{n=-N}^N x[n]x[n+m] \right) \quad (5.12)$$

in

$$S_x(F) = \text{TDFT}(R_{xx}[m]) \quad (5.13)$$

Kadar je časovno diskretni signal periodičen s periodo N , nastopajo v spektru samo mnogokratniki osnovne frekvence $F_0 = 1/N$. Spekter periodičnega časovno diskretnega signala je potem enak:

$$X_k = X(kF_0) = \frac{1}{N} \sum_{n=0}^{N-1} x[n]e^{-2\pi nk/N} \quad (5.14)$$

Zgornji izraz imenujemo diskretni Fourierov transform – DFT.

Čeprav je DFT Fourierov transform periodičnih signalov, se pogosto uporablja tudi pri obravnavi aperiodičnih in naključnih signalov. Pri njegovi uporabi se je potrebno zavedati, da z njegovo uporabo implicitno predpostavljamo, da je signal periodičen, torej da se zunaj N vzorcev preko katerih smo računali DFT, ponavlja.

Uporabnost DFT je velika tudi zato, ker obstajajo algoritmi, ki jih poznamo pod skupnim imenom hitri Fourierov transform (FFT – Fast Fourier Transform), in zmanjšajo število potrebnih računskih operacij za izračun DFT. Dolžina N mora biti pri teh algoritmihi običajno celoštevilčna potenca števila 2:

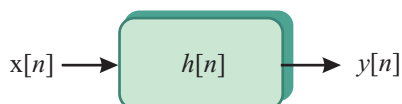
$$N = 2^M \quad (5.15)$$

Faktor zmanjšanja potrebnih računskih operacij je približno enak N/M . Zmanjšanje je zelo občutno pri velikih N oziroma velikih M . Pri $M = 10$ je faktor približno 102 pri $M = 15$ pa že več kot dva tisoč.

Digitalna sita

V poglavju so na kratko digitalna sita. Več pozornosti je posvečeno sitom s končnim odzivom, ker so ta primernejša za adaptivne sisteme. Predstavljena so adaptivna sita in osnovni algoritmi za njihovo adaptacijo. Predstavitev je omejena samo na lastnosti, ki so potrebne za razumevanje v naslednjih poglavjih.

Digitalna sita so linearni časovno nespremenljivi digitalni sistemi. Digitalno sito preslika digitalen signal $x[n]$ na svojem vhodu v digitalen signal $y[n]$ na svojem izhodu, kot je to prikazano na sliki 6.1. Podobno kot pri analognih sitih je digitalno sito popolnoma določeno s sistemsko



Slika 6.1 – Digitalno sito preslika digitalni signal $x[n]$ na svojem vhodu v digitalni signal svojemu na izhodu. Digitalno sito je popolnoma določeno s sistemsko funkcijo $h[n]$.

funkcijo $h[n]$, ki je definirana kot odziv sita na digitalen enotni impulz $\delta[n]$:

$$\delta[n] = \begin{cases} 1 & ; \quad n = 0 \\ 0 & ; \quad n \neq 0 \end{cases} \quad (6.1)$$

Izhodni signal je potem enak konvoluciji vhodnega signala in sistemske funkcije:

$$y[n] = h[n] * x[n] = \sum_{m=-\infty}^{\infty} h[m]x[n-m] \quad (6.2)$$

Sistemska funkcija kavalnega sita je enaka 0 za vse $n < 0$. Za kavalna sita je spodnja meja vsote v zgornjem izrazu enaka $m = 0$.

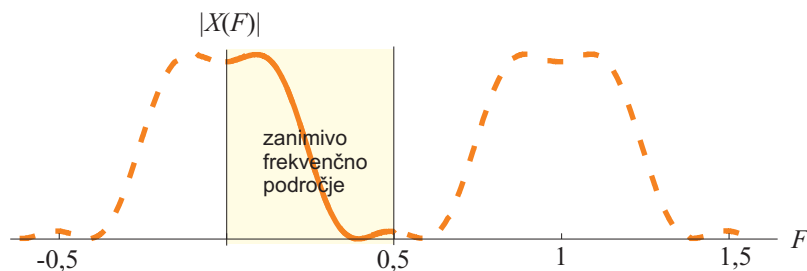
Prevalna funkcija digitalnega sita je definirana kot TDFFT sistemske funkcije:

$$H(F) = \sum_{n=-\infty}^{\infty} h[n] e^{-j2\pi Fn} \quad (6.3)$$

Omeniti velja, da je prevalna funkcija $H(F)$ periodična po frekvenci F s periodo enako 1. Zanimivo frekvenčno področje digitalnega sistema je na intervalu $0 \leq F \leq 0,5$, potem pa se karakteristika zrcali in ponavlja, kot je to prikazano na sliki 6.2.

6.1 Izvedba digitalnih sit

Ker delajo digitalna sita z digitalnimi signali, to je s števili, predstavlja njihova izvedba izvedbo določenih računskih operacij. Računske operacije lahko izvedemo neposredno z ustrezno strojno opremo ali pa z ustrezno programsko opremo na računalniku.



Slika 6.2 – Prevaljalna funkcija časovno diskretnega sistema je periodična po frekvenci F s periodo enako 1, ki ustreza vzorčni frekvenci signala. Zanimivo frekvenčno področje diskretnih sit je na intervalu $0 \leq F \leq 0,5$, potem pa se začne karakteristika zrcaliti in ponavljati.

Če je $x[n]$ vhodni signal digitalnega sita, računamo izhodni signal po enačbi:

$$y[n] = \sum_{m=0}^N b_m x[n-m] - \sum_{m=1}^M a_m y[n-m] \quad (6.4)$$

kjer so b_m in a_m uteži sita. Z utežmi b_m utežimo trenutni in pretekle vzorce vhodnega signala $x[n-m]$ v prvi vsoti z utežmi a_m pa pretekle vzorce izhodnega signala $y[n-m]$ v drugi vsoti. Ker trenutni izhodni vzorec ni odvisen samo od vhodnih vzorcev temveč tudi od preteklih izhodnih vzorcev, je enačba rekurzivna. Rekurzivni del enačbe predstavlja povratno vezavo, zaradi katere lahko postane sito nestabilno.

Sita, pri katerih so vsi koeficienti a_m enaki 0, in je trenutni izhodni vzorec $y[n]$ odvisen samo od trenutnega in preteklih vhodnih vzorcev, imenujemo nerekurzivna sita. Taka sita imajo končno dolžino odziva na enotin impulz in jih zato imenujemo tudi sita s končnim odzivom (FIR – Finite Impulse Response). FIR sita so vedno stabilna, ne glede na izbiro koeficientov b_m . Zaradi te lastnosti so zelo uporabna v adaptivnih sistemih, ki jih bomo obravnavali v nadaljevanju.

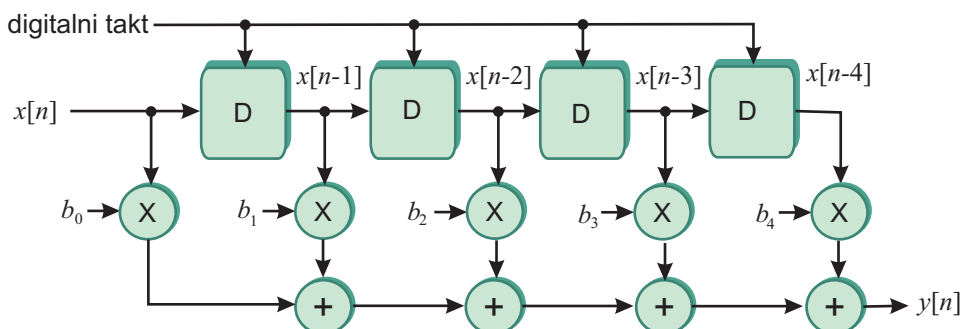
Ker se bomo v nadaljevanju srečali samo s FIR siti, si ta oglejmo nekoliko natančneje. FIR sito opisuje nerekurzivna enačba:

$$y[n] = \sum_{m=0}^N b_m x[n-m] \quad (6.5)$$

Število N , ki pove koliko preteklih vhodnih vzorcev vpliva na trenutni izhodni vzorec, imenujemo red sita. Število koeficientov b_m je enako $N + 1$, kar imenujemo tudi dolžina FIR sita. Strojna izvedba FIR sita z množilniki, seštevalniki in zakasnilnimi D celicami je prikazana na slik 6.3.

Vhodni signal $x[n]$ se začasno shrani v D celici in se ob vsakem digitalnem taktu pomakne v naslednjo celico, tako, da so v zakasnilnih celicah shranjeni pretekli vzorci signala. Tako verigo zakasnilnih celic imenujemo tudi pomikalni register. Z množilniki množimo vzorca shranjene v D celicah z utežmi b_m in jih nato s seštevalniki seštejemo, da dobimo izhodni signal.

Če pripeljemo na vhod takega sita enotin impulz $\delta[n]$, dobimo na izhodu sistemsko funkcijo $h[n]$. Enotin impulz je ima vrednost 1 pri $n = 0$, vsi ostali vzorci pa so enaki 0. V času $n = 0$ dobimo torej na izhodu samo vrednost b_0 , potem se ob digitalne taktu vrednost 1 pomakne v



Slika 6.3 – Strojna izvedba fir sita 4. reda s seštevalniki množilniki in zakasnilnimi D celicami.

naslednjo celico, in na izhodu dobimo vrednost b_1 . Proces se nadaljuje, dokler 1 ne pride do zadnje D celice in dobimo na izhodu vrednost b_N (v primeru na sliki b_4), pomikalni register pa se ob naslednjem taktu izprazni. Vidimo, da je sistemska funkcija FIR sita neposredno določena s koeficienti b_m :

$$h[n] = b_n ; 0 \leq n \leq N \quad (6.6)$$

Ker poznamo sistemsko funkcijo FIR sita, lahko izračunamo z TDF^T izračunamo tudi njegovo prevažalno funkcijo $H(F)$:

$$H(F) = \sum_{n=0}^N h[n] e^{-j2\pi Fn} = \sum_{n=0}^N b_n e^{-j2\pi Fn} \quad (6.7)$$

Enačbo FIR sita (6.5) lahko zaradi boljše preglednosti zapišemo tudi v vektorski obliki. Definiramo stolpični vektor uteži:

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \dots \\ b_N \end{bmatrix} \quad (6.8)$$

in stolpični vektor zadnjega in N preteklih vzorcev signala, ki so shranjeni v pomikalnem registru:

$$\mathbf{x}[n] = \begin{bmatrix} x[n] \\ x[n-1] \\ x[n-2] \\ \dots \\ x[n-N] \end{bmatrix} \quad (6.9)$$

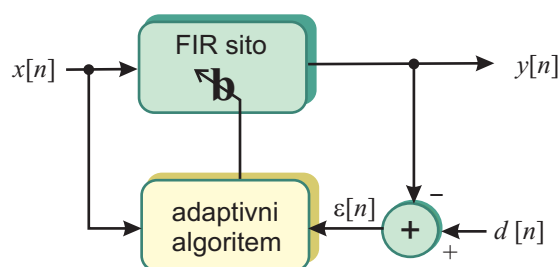
Izhodni signal FIR lahko potem izrazim kot skalarni produkt obeh vektorjev:

$$y[n] = \mathbf{b}^T \cdot \mathbf{x}[n] \quad (6.10)$$

6.2 Adaptivna sita

Adaptivna sita so sita, ki so sposobna med delovanjem prilagajati svojo karakteristiko okolju v katerem delujejo, oziroma signalu na svojem vhodu. Karakteristika digitalnega sita je odvisna od uteži sita, tako da lahko s spreminjanjem uteži, spreminjamo njegovo karakteristiko. Pri spremembah uteži lahko postane sito nestabilno, zato so za adaptivne sisteme primernejša FIR sita, ki ne morejo postati nestabilna, ne glede na izbiro uteži.

Načelna vezava adaptivnega sita je prikazana na sliki 6.4. Na vhod sita pripeljemo signal



Slika 6.4 – Adaptivno sito. Uteži sita \mathbf{b} prilagajamo tako, da je signal na izhodu sita $y[n]$ čim bolj podoben referenčnemu signalu $d[n]$. Uteži prilagajamo z ustreznim algoritmom, ki minimizira srednjo kvadratično vrednost napake $\epsilon[n]$.

$x[n]$. Želimo, da bi bil izhodni signal $y[n]$ čim bolj podoben referenčnemu signalu $d[n]$. Njunjo razliko imenujemo zato signal napake $\epsilon[n]$. Velja:

$$\epsilon[n] = d[n] - y[n] = d[n] - \mathbf{b}^T \cdot \mathbf{x}[n] \quad (6.11)$$

Kot kriterijsko funkcijo podobnosti ξ vzamemo moč signala napake, ki je enaka srednji kvadratični vrednosti napake:

$$\xi = \overline{\epsilon^2[n]} = \overline{d^2[n]} - 2\mathbf{b}^T \cdot \mathbf{p} + \mathbf{b}^T \cdot \mathbf{R} \cdot \mathbf{b} \quad (6.12)$$

kjer je \mathbf{p} korelacijski vektor vhodnega in referenčnega signala:

$$\mathbf{p} = \overline{\mathbf{x}[n] d[n]} \quad (6.13)$$

in \mathbf{R} avtokorelacijska matrika vhodnega signala:

$$\mathbf{R} = \overline{\mathbf{x}[n] \cdot \mathbf{x}^T[n]} \quad (6.14)$$

Srednjo kvadratično vrednost napake ξ vzamemo za kriterijsko funkcijo, ki jo želimo minimizirati. Ker je ξ kvadratna funkcija uteži \mathbf{b} , kar pomeni, ima en sam globalni minimum. Gradient $\nabla \xi$ je vektor, ki kaže v smeri maksimalnega naklona kriterijske funkcije, in ga dobimo

z odvajanjem ξ po vseh utežeh:

$$\nabla \xi = \frac{\partial \xi}{\partial \mathbf{b}} = \begin{bmatrix} \frac{\partial \xi}{\partial b_1} \\ \frac{\partial \xi}{\partial b_2} \\ \frac{\partial \xi}{\partial b_3} \\ \dots \\ \frac{\partial \xi}{\partial b_N} \end{bmatrix} \quad (6.15)$$

Kriterijska funkcija bo minimalna, kadar je gradient enak $\mathbf{0}$:

$$\nabla \xi = 2(\mathbf{R} \cdot \mathbf{b} - \mathbf{p}) = \mathbf{0} \quad (6.16)$$

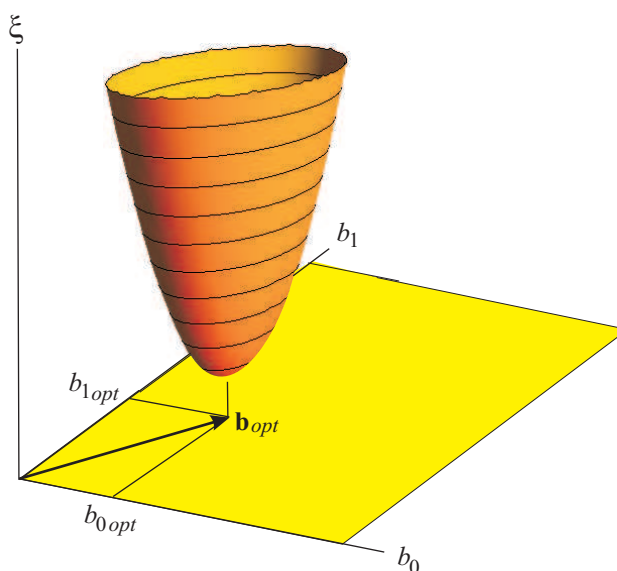
Rešitev te enačbe na \mathbf{b} da optimalen vektor uteži:

$$\mathbf{b}_{opt} = \mathbf{R}^{-1} \cdot \mathbf{p} \quad (6.17)$$

Če bi poznali avtokorelacijsko matriko \mathbf{R} in korelacijski vektor \mathbf{p} , bi lahko izračunali optimalno vrednost vektorja uteži \mathbf{b} . Adaptivno sito potrebujemo takrat, kadar ju ne poznamo vnaprej oziroma se spreminjata med delovanjem.

6.2.1 Gradientni algoritem

Kot smo že omenili ima kriterijska funkcija ξ en sam globalni minimum, kot je to za primer FIR sita z dvema utežmi prikazano na sliki 6.5.



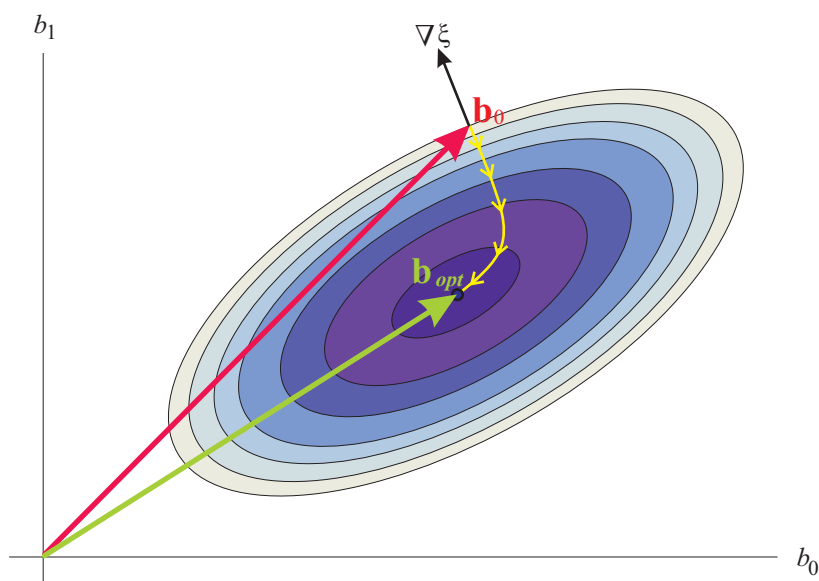
Slika 6.5 – Primer kriterijske funkcije FIR sita z dvema utežmi b_0 in b_1 . Kriterijska funkcija ima globalni minimum pri optimalni vrednosti vektorja \mathbf{b} .

Pri gradientnem algoritmu si izberemo neko začetno utež \mathbf{b}_0 in jo nato iterativno popravljamo tako, da se spuščamo po največji strmini kriterijske funkcije – tako kot bi tekla voda. Ker

gradient kaže v smeri največje strmine navzgor, moramo uteži popravljati v nasprotni smeri gradienta, po iterativni enačbi:

$$\mathbf{b}_{m+1} = \mathbf{b}_m - \mu \nabla \xi \quad (6.18)$$

Na sliki 6.6 je prikazan primer kriterijske funkcije za FIR sito z dvema utežma v pogledu z vrha navzdol, tako da so vidne izohipse, to je črte ki povezujejo točke z enako vrednostjo kriterijske funkcije. Optimalni vrednosti vektorja uteži \mathbf{b}_{opt} se približujemo po krivulji, ki je ves čas



Slika 6.6 – Primer kriterijske funkcije FIR sita z dvema utežma b_0 in b_1 – pogled z vrha navzdol. Izohipse so pri dveh utežeh elipse. Gradient v vsaki točki je pravokoten na izohipso, ki gre skozi to točko. Od začetne vrednosti vektorja uteži \mathbf{b}_0 popravljamo uteži v nasprotni smeri gradienta po krivulji, ki je pravokotna na izohipse, proti optimalni vrednosti \mathbf{b}_{opt} .

pravokotna na izohipse. Adaptacijski faktor μ določa velikost adaptacijskega koraka in s tem hitrost adaptacije. Prevelik adaptacijski faktor lahko povzroči, da se z vsakim korakom še bolj oddaljimo od optimalne vrednosti. Postopek postane divergenten.

Da bi lahko popravljali uteži v nasprotni smeri gradienta $\nabla \xi$, moramo gradient najprej določiti. V skladu z izrazom (6.15) moramo pridobiti parcialne odvode kriterijske funkcije ξ po vseh utežeh. Odvod za k -to utež lahko določimo tako, da utež najprej povečamo za nek Δb , izmerimo $\xi(b_k + \Delta b)$, nato jo zmanjšamo za Δb in izmerimo $\xi(b_k - \Delta b)$. Odvod je potem enak:

$$\frac{\partial \xi}{\partial b_k} = \frac{\xi(b_k + \Delta b) - \xi(b_k - \Delta b)}{2\Delta b} \quad (6.19)$$

Tak postopek imenujemo perturbacija uteži. Problem je v tem, da je kriterijska funkcija ξ srednja kvadratična vrednost napake. Da bi lahko izračunali srednjo vrednost moramo napako $\epsilon[n]$ pomeriti za veliko, recimo M vzorcev. Ker rabimo v skladu z zgornjo enačbo za izračun gradienta za vsako utež po dve vrednosti ξ in je vseh uteži N , moramo za izračun gradienta napako pomeriti za $2NM$ vzorcev signala, šele potem lahko popravimo uteži. To lahko zelo upočasnjuje postopek adaptacije.

Navedimo sedaj nekaj osnovnih lastnosti gradientnega postopka s perturbacijo uteži, ne da bi jih natančno matematično utemeljevali:

- S številom vzorcev M , ki jih upoštevamo pri izračunu ξ povečujemo natančnost izračuna gradientna. Večji M pomeni natančnejši izračun gradienta vendar tudi daljši čas računanja in s tem počasnejšo adaptacijo.
- Natančno izračunan gradient je zelo pomemben v stacionarnem stanju, ko smo blizu optimalne vrednosti uteži. Ko je vrednost uteži optimalna, bi moral biti gradient enak nič, in uteži se ne bi več spreminjale. Če izračun gradienta ni natančen, dobimo neko vrednost različno od nič, in uteži popravljamo stran od optimalne vrednosti. To poveča šum v signalu.
- Zaradi perturbacije uteži spreminjamo uteži za $\pm\Delta b$ tudi, ko smo v optimalni vrednosti. Tudi to povečuje šum v signalu. Temu se lahko izognemo tako, da gradient računamo v paralelnem sistemu, kjer spreminjamo uteži, v delujočem sistemu pa jih ne perturbiramo temveč jih samo nastavljamo, ko je gradient izračunan.
- Z adaptacijskim faktorjem μ kontroliramo hitrost adaptacije. Kadar je gradient izračunan nenatančno, moramo imeti manjši μ , da v stacionarnem stanju ne bi bili napačni popravki uteži preveliki. Če zmanjšamo M zmanjšamo čas za izračun gradienta, s čimer bi lahko pospešili adaptacijo, vendar računamo gradient manj natančno, zato moramo zmanjšati μ , s čimer pa zopet upočasnimo adaptacijo, tako da ostane hitrost adaptacije praktično nespremenjena.
- Pri gradientnem postopku uteži ne popravljamo v smeri optimalne vrednosti, temveč se tej vrednosti približujemo po neki krivulji, ki je lahko pri velikem številu uteži precej daljša od direktne poti. Na podoben način kot tudi reka ne teče proti morju po najkrajši poti. To lahko precej podaljša čas adaptacije.
- Obstajajo veliko število drugih algoritmov, ki so računsko zahtevnejši, a omogočajo hitrejšo adaptacijo. Newtonov algoritem, na primer, popravlja uteži točno v smeri optimalne vrednosti in omogoča bistveno hitrejšo adaptacijo, vendar je računsko precej zahtevnejši, kar pomeni, da potrebujemo bistveno zmoglivejši procesor, za njegovo izvedbo. Opis Newtonovega algoritma presega okvirje te knjige.

6.2.2 LMS algoritem

LMS (Least Mean Square – najmanjša kvadratična srednja vrednost) algoritem je izpeljan iz gradientnega algoritma. Njegova največja prednost je v njegovi računski nezahtevnosti.

V prejšnjem razdelku smo ugotovili, da ne spremenimo bistveno šuma v stacionarnem stanju niti hitrosti adaptacije, če zmanjšamo, če zmanjšamo čas povprečenja M pri izračunu gradienta in hkrati zmanjšamo adaptacijski faktor μ . V skrajnem primeru lahko zmanjšamo M na vrednost

1, kar pomeni, da povprečimo napako samo čez en vzorec, oziroma, da namesto povprečne vrednosti vzamemo kar trenutno vrednost kvadrata napake:

$$\xi_n = \epsilon^2[n] \quad (6.20)$$

Gradient lahko tako izračunamo iz trenutne vrednosti napake v vsakem vzorčnem trenutku:

$$\nabla \xi_n = \frac{\partial \epsilon^2[n]}{\partial \mathbf{b}} = 2\epsilon[n] \frac{\partial \epsilon[n]}{\partial \mathbf{b}} \quad (6.21)$$

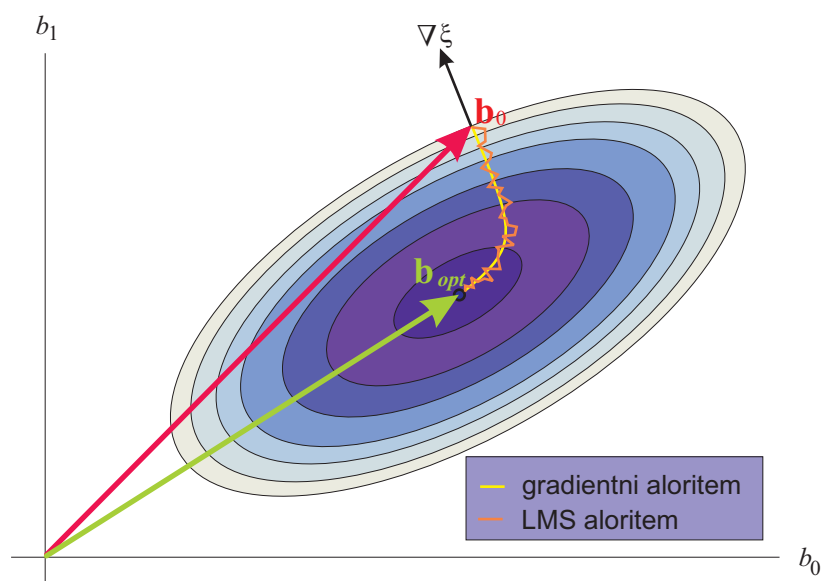
Če v zgornjo enačbo vstavimo $\epsilon[n]$ iz izraza (6.11) in odvajamo dobimo:

$$\nabla \xi_n = -2\epsilon[n]\mathbf{x}[n] \quad (6.22)$$

Ker računamo gradient po zgornji enačbi v vsakem časovnem trenutku, lahko tudi v vsakem časovnem trenutku popravljamo uteži v nasprotni smeri gradienta:

$$\mathbf{b}_{n+1} = \mathbf{b}_n - \mu \nabla \xi_n = \mathbf{b}_n + 2\epsilon[n]\mathbf{x}[n] \quad (6.23)$$

Zgornji izraz predstavlja LMS algoritem za popravljanje uteži. Ker smo namesto srednje kvadratične vrednosti napake vzeli kar kvadrat trenutne napake, je lahko izračun gradienta zelo napačen, zato moramo močno zmanjšati adaptacijski faktor μ . Ker so napake v izračunu gradienta naključne in vedno v različno smer, se med seboj izničijo in se uteži še vedno spreminjajo proti optimalni vrednosti v okolici krivulje, ki je v smeri pravega gradienta, kot je to prikazano na sliki 6.7.



Slika 6.7 – Pri LMS algoritmu se uteži spreminjajo proti optimalni vrednosti po približni krivulji gradientnega postopka. Napake se zaradi različnih smeri napak izničijo. V bistvu se namesto povprečenja srednje vrednosti tu povprečijo gradienti.

Ker popravljamo uteži v vsakem časovnem trenutku, se izkaže, da je pri istem šumu v stacionarnem stanju, kljub potrebnemu zmanjšanju adaptacijskega faktorja μ , adaptacija pri LMS algoritmu hitrejša kot adaptacija pri gradientnem postopku s perturbacijo uteži.

Tudi LMS algoritem se da na račun povečane računske zahtevnosti predelati tako, da so popravki direktno v smeri optimuma in ne v smeri gradienta.

Digitalen prenos v osnovnem pasu

Poglavje obravnava prenos niza simbolov oziroma niza števil v osnovnem pasu. Opisano je prenos z impulzi v osnovnem pasu. Definirana je intersimbolna interferenca in podan Nyquistov kriterij za prenos brez intersimbolne interference. Določena je minimalna pasovna širina, ki je potrebna za prenos simbolov v osnovnem pasu brez intersimbolne interference. Opisani so tudi različni postopki za izločanje intersimbolne interference.

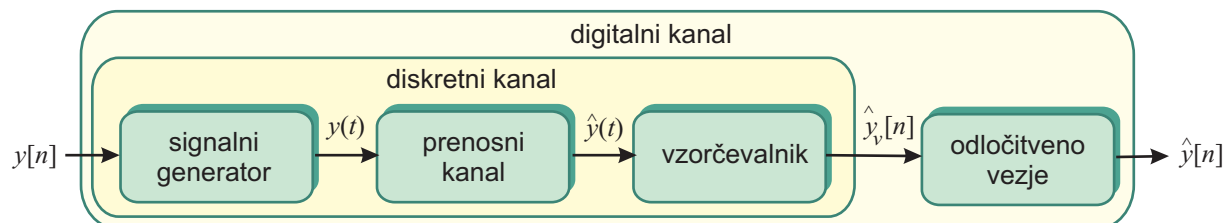
Za razliko od analognega prenosa pri digitalnem prenosu ne gre za prenos signalne oblike, temveč za prenos niza simbolov. Niz poljubnih simbolov (na primer črk v tekstu) pred prenosom pretvorimo v niz števil oziroma v digitalen signal. Digitalno lahko prenašamo podatke med računalniki (podatkovni prenos) ali pa vzorce analognih signalov, kot je to primer pri digitalnem prenosu govornega signala v digitalni telefoniji.

Označimo z $x[n]$ digitalen signal. Števila v nizu imajo omejen nabor vrednosti – omejeno abecedo. Zavzamejo lahko le vrednosti iz abecede, kot na primer:

$$\begin{aligned} \text{binaren niz} \quad x[n] &\in \{0, 1\} \\ \text{bipolaren niz} \quad x[n] &\in \{-1, 1\} \\ \text{štirinivojski niz} \quad x[n] &\in \{-3, -1, 1, 3\} \end{aligned}$$

Ker imajo omejeno abecedo imenujemo ta števila tudi številski simboli.

Splošen model nekodiranega digitalnega prenosa je prikazan na sliki 7.1. Ta predstavlja del osnovnega komunikacijskega modela s slike 1.1, to je del od izhoda kanalskega kodirnika do vhoda kanalskega dekodirnika.



Slika 7.1 – Splošen model nekodiranega digitalnega prenosa. Prenosni kanal skupaj s signalnim generatorjem in vzorčevalnikom tvorijo tako imenovani diskretni (nezvezen po času) kanal, ki preslika niz simbolov na svojem vходу v niz simbolov na izhodu. Simboli na izhodu so realna števila.. Če dodamo diskretnemu kanalu še odločitveno vezje dobimo digitalen kanal, ki preslika niz simbolov $y[n]$ v niz sprejetih simbolov $\hat{y}[n]$, ki predstavljajo odločitve o oddanih simbolih, in želimo, da so enaki oddanim simbolom $y[n]$.

Niz simbolov $y[n]$ s signalnim generatorjem pretvorimo v signal $y(t)$, ki je primeren za izbran prenosni kanal. V prenosnem kanalu se signal $y(t)$ popači, poleg tega pa se mu doda šum oziroma motnje, tako, da na izhodu signala dobimo približek tega signala $\hat{y}(t)$. Signal $y(t)$ nato vzorčimo in dobimo niz vzorcev $\hat{y}_v[n]$. Signalni generator, prenosni kanal in vzorčevalnik tvorijo diskreten kanal. Zaradi popačenja in šuma v prenosnem kanalu so vzorci na izhodu diskretnega kanala realna števila, ki ne pripadajo več naboru vrednosti oddanih simbolov.

Odločitveno vezje se na osnovi vrednosti vzorcev, ki jih dobi iz diskretnega kanala, odloča, kateri simboli iz abecede so bili oddani. Na osnovi teh odločitev tvori sprejeti niz $\hat{y}[n]$. Diskreten kanal skupaj z odločitvenim vezjem tvori digitalen kanal, ki preslika niz simbolov $y[n]$ na svojem vhodu, v niz simbolov $\hat{y}[n]$ na svojem izhodu.

Medtem, ko pri analognem prenosu želimo, da je signal $\hat{y}(t)$ čim bolj podoben signalu $y(t)$, pa želimo pri digitalnem prenosu, da je niz $\hat{y}[n]$ čim bolj podoben nizu $y[n]$. Pri prenosu brez napak sta ta dva niza enaka. Kvaliteto digitalnega kanala ocenjujemo s pogostostjo simbolnih napak (SER – Symbol Error Rate):

$$SER = \frac{N_n}{N} \quad (7.1)$$

kjer je N_n število napačno sprejetih simbolov in N število vseh oddanih simbolov. Kadar gre za binarni ali bipolarni niz je SER enaka pogostosti bitnih napak (BER – Bit Error Rate). Pri večnivojskem prenosu se ti dve vrednosti lahko razlikujeta, o čemer bo več govora kasneje.

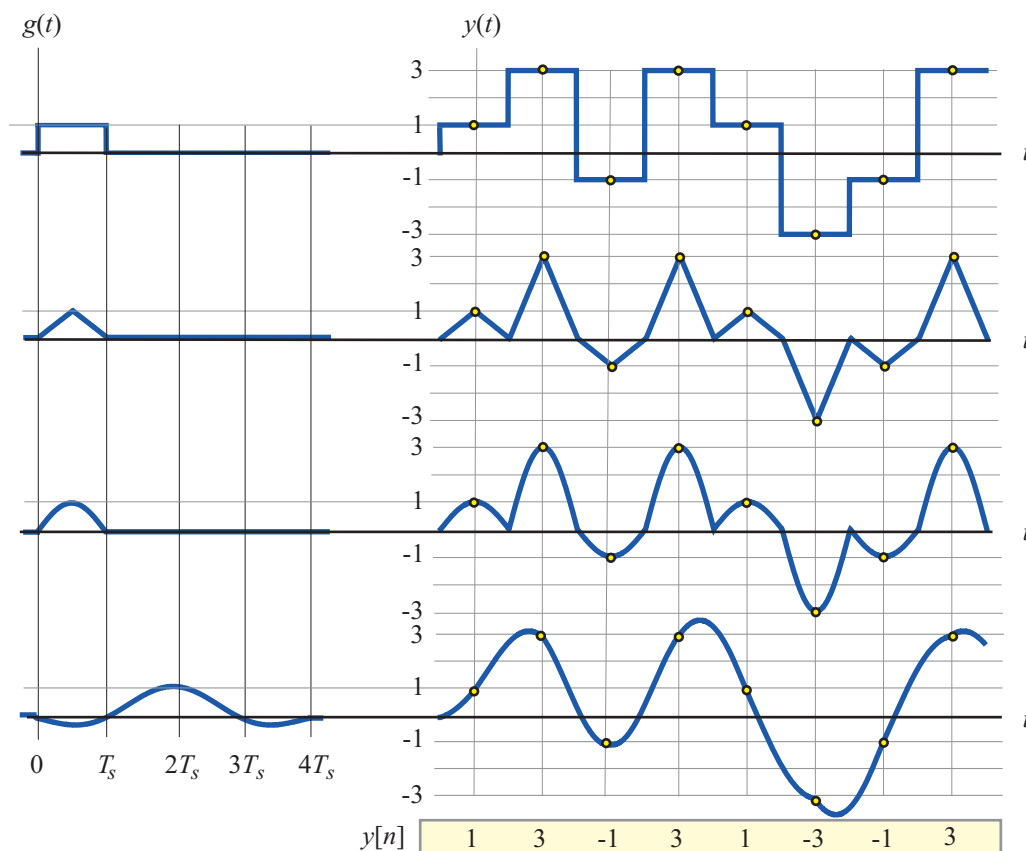
Digitalen prenos v osnovnem pasu večinoma izvajamo s prenosom impulzov v osnovnem pasu. Signalni generator na sliki 7.1 je v tem primeru generator impulzov. Simboli niza $y[n]$ prihajajo na signalni generator s simbolno hitrostjo f_s :

$$f_s = \frac{1}{T_s} \quad (7.2)$$

kjer je s T_s označen čas med dvema simboloma. Signalni generator na svojem izhodu tvori impulze z amplitudami enakimi vrednostim pripadajočih simbolov $y[n]$ na svojem vhodu. Oblika impulzov je lahko poljubna, zelo pogosto pa se uporabljajo pravokotni impulzi dolžine T_s . Signal $y(t)$ lahko zapišemo kot vsoto časovno zamaknjenih impulzov. Če z $g(t)$ označimo obliko impulza, lahko zapišemo:

$$y(t) = \sum_{n=0}^{N-1} y[n] g(t - nT_s) \quad (7.3)$$

Primeri signala na izhodu signalnega generatorja pri različnih oblikah impulzov so prikazani na sliki 7.2. S krožci so označene vrednosti, ki bi jih dobili v vzorčnih trenutkih, če na prenosnem kanalu ne bi prišlo do popačenja signala in če ne bi bilo šuma. Zadnji impulz sega čez več simbolnih intervalov T_s , zato se prikrite vrednosti zamaknjenih impulzov seštevajo. Ker gre impulz ob mnogokratnikih T_s skozi nič, to ne spremeni vrednosti signala $y(t)$ ob trenutkih vzorčenja, tako da imajo vzorci še vedno pravilne vrednosti.



Slika 7.2 – Primeri impulzov pri prenosu v osnovnem pasu. Impulzi so v splošnem lahko tudi daljši od simbolnega časa T_s . V tem primeru pride do prekrivanja impulzov, vendar so lahko vrednosti v trenutkih, ko bomo vzorčili signal še vedno pravilne.

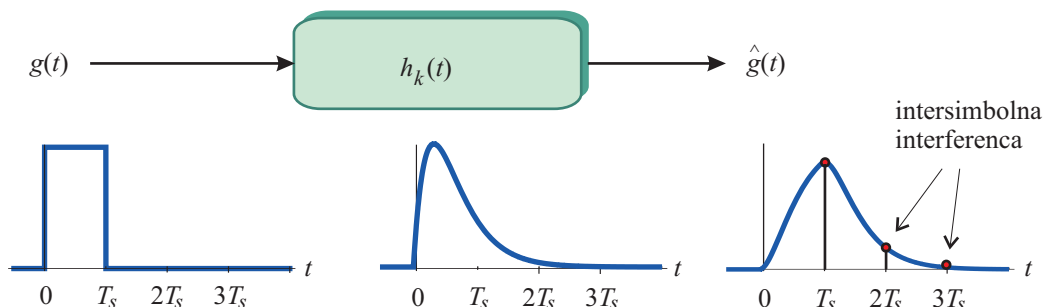
7.1 Intersimbolna interferenca

Če bi imeli na voljo idealen prenosni kanal, bi bil signal $\hat{y}(t)$ enak oddanemu signalu $y(t)$, njegovi vzorci $\hat{y}_v[n]$ pa enaki oddanim simbolom $y[n]$. Zaradi šuma in popačenja v prenosnem kanalu, pa se vzorci signala razlikujejo od oddanih simbolov. Predpostavimo, da je prenosni kanal linearen in časovno nespremenljiv. Prenosni kanal lahko potem opišemo s sistemsko funkcijo $h_k(t)$. Pri prehodu skozi linearen sistem se impulzi ki jih oddajamo oslabijo raztegnejo po času, saj velja:

$$\hat{g}(t) = g(t) * h_k(t) \quad (7.4)$$

kot je to za primer pravokotnega impulza prikazano na sliki 7.3.

Ker pripeljemo v skladu z izrazom (7.3) na vhod prenosnega kanala vsoto impulzov, dobimo na izhodu prenosnega kanala vsoto njihovih odzivov. Pri sprejemu impulze vzorčimo pri njihovi maksimalni vrednosti. Kadar se impulz razteza v čas ko vzorčimo naslednji impulz, se ta vrednost prišteje naslednjemu vzorcu in spremeni njegovo vrednost. Ta pojav imenujemo intersimbolna interferenca.



Slika 7.3 – Popačenje impulza pri prenosu. Impulz se raztegne po časovni osi. Pri sprejemu bi tak impulz vzorčili pri njegovi maksimalni vrednosti. Problem je, da se razteza tudi v čas, ko bi vzorčili naslednji impulz in pokvari njegov vzorec. Pride do medsebojnega vpliva med oddanimi simboli ali intersimbolne interference

7.2 Nyquistov kriterij za prenos brez intersimbolne interference

Videli smo, da je intersimbolna interferenca odvisna tako od oblike impulza kot tudi od karakteristike prenosnega kanala. Da bi poenostavili analizo, bomo signalni generator razdelili na dva dela, na generator ki generira enotne impulze $\delta(t - nT_s)$ in sito za oblikovanje impulzov, ki enotne impulze preoblikuje v impulze želene oblike $g(t - nT_s)$.

Spomnimo se sedaj, da je sistemska funkcija linearnega časovno nespremenljivega sistema definirana kot odziv sistema na enotin impulz. Ker so na vhodu sita za oblikovanje enotini impulzi, na njegovem izhodu pa pričakujemo impulze oblike $g(t)$, mora biti sistemska funkcija sita za oblikovanje kar enaka obliki impulza $g(t)$.

To, da smo razdelili signalni generator na dva dela, nikakor ne pomeni, da bo njegova izvedba v praksi takšna. Za generacijo pravokotnih impulzov seveda ni potrebno generirati enotnih impulzov in jih potem preoblikovati temveč jih mnogo lažje generiramo neposredno. Celo več, v praksi sploh ni mogoče generirati enotnih impulzov saj imajo neskončno amplitudo in so neskončno ozki. Je pa tak pristop primeren za analizo, saj imamo sedaj na vhodu vedno enake impulze, karakteristiko sita za oblikovanje pa lahko pridružimo karakteristiki prenosnega kanala.

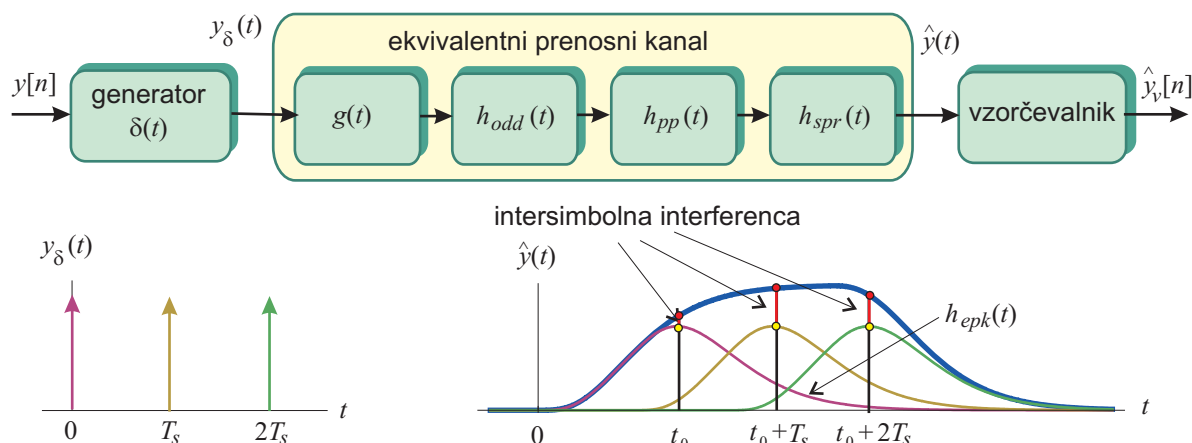
Prenosni kanal poleg prenosne poti s sistemsko funkcijo $h_{pp}(t)$ vsebuje še oddajnik in sprejemnik. Oddajnik in sprejemnik imata tudi svojo frekvenčno karakteristiko, ki ravno tako vpliva na popačenje impulzov in s tem na intersimbolno interferenco. Karakteristiki oddajnika in sprejemnika pri prenosu v osnovnem pasu lahko ponazorimo z oddajnim sitom in sprejemnim sitom s sistemskima funkcijama $h_{odd}(t)$ in $h_{spr}(t)$.

Z upoštevanjem vseh zgornjih predpostavk, dobimo model sistema za prenos impulzov v osnovnem pasu, ki je prikazan na sliki 7.4

Ker so sito za oblikovanje, oddajno sito, prenosna pot in sprejemno sito vezani v serijo, jih lahko nadomestimo z ekvivalentnim prenosnim kanalom s sistemsko funkcijo:

$$h_{epk}(t) = g(t) * h_{odd}(t) * h_{pp}(t) * h_{spr}(t) \quad (7.5)$$

Na vhodu ekvivalentnega prenosnega kanal imamo sedaj niz enotnih impulzov uteženih z simboli $y[n]$. Vsak od teh impulzov povzroči na izhodu ekvivalentnega prenosnega kanala odziv, ki je



Slika 7.4 – Prenos impulzov v osnovnem pasu. Sito za oblikovanje, oddajno sito, prenosna pot in sprejemno sito so vezani v serijo in skupaj tvorijo ekvivalentni prenosni kanal s sistemsko funkcijo $h_{epk}(t)$.

enak sistemski funkciji tega kanala. Vsi ti odzivi so med seboj sešteti. Signal vzorčimo s simbolno hitrostjo f_s in z zakasnitvijo t_0 , ki ustreza zakasnitvi na kanalu. Če se odziv enega simbola razteza čez vzorčni trenutek drugega simbola, povzroča intersimbolno interferenco. Na sliki 7.4 so rdeče označeni vzorci signala, ki jih dobimo iz vzorčevalnika, rumeno pa vzorci, ki bi jih dobili, če ne bi bilo intersimbolne interference.

Na sliki 7.4 lahko vidimo, da do intersimbolne interference ne bi prišlo, če bi imel odziv na enotin impulz vrednost 0 v vseh vzorčnih trenutkih razen v trenutku, ko vzorčimo zelen simbol. Do podobne ugotovitve smo prišli že pri zadnjem impulzu na sliki 7.2, ki se razteza preko več simbolnih intervalov.

7.2.1 Nyquistov kriterij v časovnem prostoru

Na osnovi zgornjega razmišljanja lahko zapišemo pogoj za prenos brez intersimbolne interference kot:

$$h_{epk}(t_0 + nT_s) = \begin{cases} A & ; n = 0 \\ 0 & ; n \neq 0 \end{cases} \quad (7.6)$$

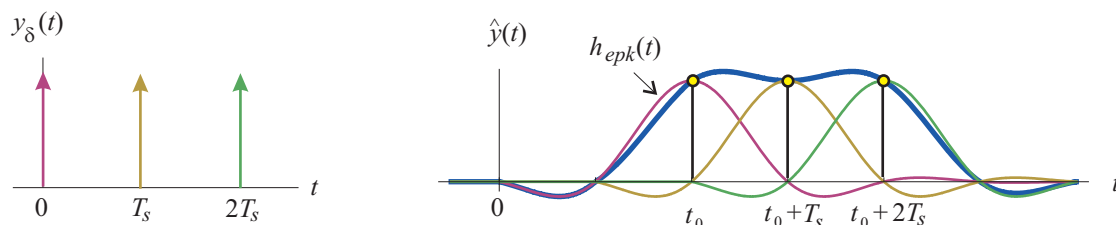
Ta pogoj imenujemo Nyquistov kriterij za prenos brez intersimbolne interference v časovnem prostoru. Vrednost A je v realnih sistemih običajno manjša od 1 in predstavlja slabljenje sistema, ki pa ni problematično, saj lahko vključimo ustrezen ojačevalnik, oziroma lahko sprejete vzorce množimo z inverzno vrednostjo A .

Brez izgube splošnosti lahko postavimo zakasnitev $t_0 = 0$, saj za pravilno nastavitve časa vzorčenja v praksi skrbi sinhronizacijsko vezje. Ob tej predpostavki se Nyquistov kriterij poenostavi v:

$$h_{epk}(nT_s) = \begin{cases} A & ; n = 0 \\ 0 & ; n \neq 0 \end{cases} \quad (7.7)$$

kar bomo zaradi preprostejših izrazov uporabljali v nadaljevanju.

Na sliki 7.5 je prikazan primer odziva na tri zaporedne enake enotine impulze, kadra sistemska funkcije ekvivalentnega prenosnega kanala izpolnjuje Nyquistov kriterij in ne pride do intersimbolne interference, kot na sliki 7.4.



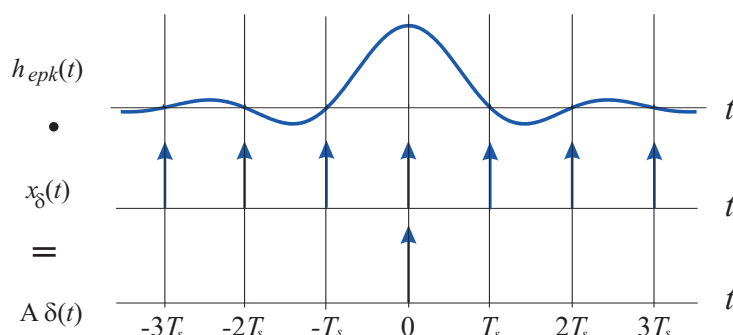
Slika 7.5 – Primer odziva na tri enake zaporedne simbole, kadar sistemska funkcija izpolnjuje Nyquistov kriterij in ni intersimbolne interference.

7.2.2 Nyquistov kriterij v frekvenčnem prostoru

Nyquistov kriterij v časovnem prostoru postavlja pogoj za sistemska funkcijo ekvivalentnega prenosnega kanala $h_{epk}(t)$. Postavi se vprašanje, kakšen je pogoj, ki ga mora izpolnjevati prevajalna funkcija ekvivalentnega prenosnega kanala $H_{epk}(f)$, da ne bi prišlo do intersimbolne interference.

Pogoj (7.7) predpisuje vrednosti $h_{epk}(t)$ zgolj ob vzorčnih trenutkih nT_s , izven teh trenutkov pa je lahko njen potek poljuben. Zato ne moremo narediti Fourierovega transformata tega izraza. Da bi lahko prišli do pogoja v frekvenčnem prostoru, moramo najprej pogoj v časovnem prostoru zapisati nekoliko drugače.

Sistemska funkcijo $h_{epk}(t)$ vzorčimo vzorčevalno funkcijo $v_\delta(t)$, kot smo jo definirali v (5.3), pri čemer postavimo da je frekvenca vzorčenja f_{vz} enaka simbolni hitrosti f_s oziroma, da velja $T_{vz} = T_s$. Če sistemska funkcija $h_{epk}(t)$ izpolnjuje Nyquistov kriterij podan z enačbo (7.7), potem je njena vrednost povsod, kjer nastopajo enotini impulzi vzorčevalne funkcije, enaka nič, razen v času $t = 0$, kjer je amplituda enaka A , kot je to prikazano na sliki 7.6. Nyquistov kriterij



Slika 7.6 – Pri množenju sistemske funkcije, ki izpolnjuje Nyquistov kriterij, z nizom enotinih impulzov, dobimo kot rezultat enotin impulz v času $t = 0$

v časovnem prostoru lahko sedaj zapišemo kot:

$$h_{epk}(t) \cdot v_\delta(t) = A \delta(t) \quad (7.8)$$

Zgornji pogoj je popolnoma ekvivalenten pogoj (eq. Nyquistov Kriterij CasBrezZakasnitve). Zapisan v tej obliki omogoča, da naredimo Fourierov transform na obeh straneh enačbe. Na enak način, kot smo v enačbi (5.6) ugotovili, da se pri vzorčenju signala, njegov spekter preslika okrog vseh mnogokratnikov vzorčne frekvence, lahko ugotovimo, da se pri vzorčenju sistemske funkcije, prevajalna funkcija preslika okrog mnogokratnikov vzorčne frekvence. Iz lastnosti Fourierovega transformata tudi vemo, da je Fourierov transform enotinega impulza enak 1. Če naredimo sedaj Fourierov transform izraza (7.8), dobimo po preureditvi:

$$\sum_{k=-\infty}^{\infty} H_{epk}(f + kf_s) = AT_s \quad (7.9)$$

Zgornji izraz imenujemo Nyquistov kriterij za prenos brez intersimbolne interference v frekvenčnem prostoru. Na levi strani enačbe nastopa vsota za mnogokratnike f_s zamaknjenih prevajalnih funkcij ekvivalentnega prenosnega kanala. Ker se premaknjene funkcije med seboj prekrivajo imenujemo to tudi *prekrita prevajalna funkcija*. Na desni strani izraza pa je konstantna vrednost AT_s . Če to izrazimo z besedami, se Nyquistov kriterij za prenos brez intersimbolne interference v frekvenčnem prostoru glasi:

Da pri digitalnem prenosu ne bi prišlo do intersimbolne interference mora biti prekrita prevajalna funkcija ekvivalentnega prenosnega kanala konstantna. V ekvivalentni prenosni kanal so vključeni vsi elementi, ki oblikujejo signal, to so sito za oblikovanje impulzov, oddajno sito, prenosna pot in sprejemno sito.

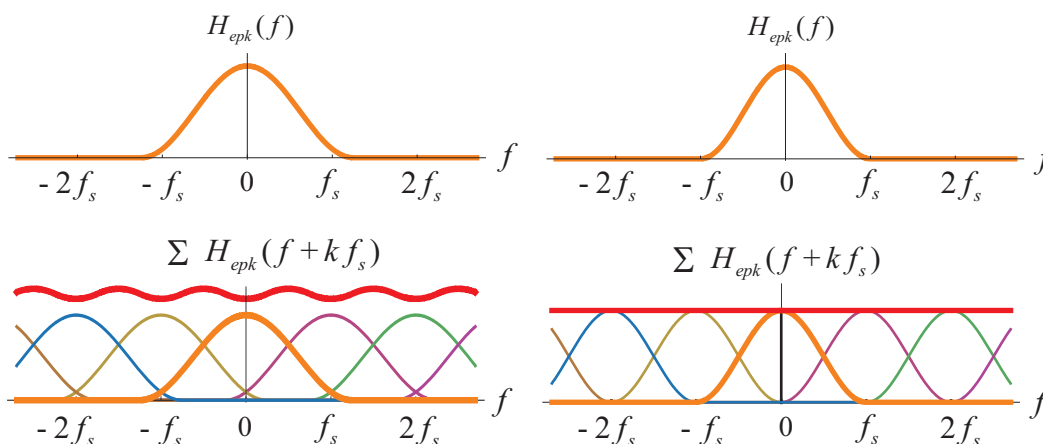
Na sliki 7.7 sta prikazana dva primera prevajalnih funkcij. Prevajalna funkcija na levi ne izpolnjuje Nyquistovega kriterija, saj prekrita prevajalna funkcija (rdeča) ni konstantna, medtem ko prevajalna funkcija na desni izpolnjuje Nyquistov kriterij. To pomeni, da pride v prvem primeru pri prenosu do intersimbolne interference, medtem, ko je prenos v drugem primeru brez intersimbolne interference.

Na levi strani slike lahko vidimo tudi, da je prekrita prevajalna funkcija periodična s periodo f_s , obenem pa je soda funkcija frekvence. Če je konstantna na intervalu med 0 in f_s , je konstantna za vse frekvence, torej zadošča, da konstantnost preverimo samo na tem intervalu, kot je to prikazano na sliki 7.8.

Ker se prevajalna funkcija navidezno prepogiba v interval med 0 in f_s jo imenujemo tudi prepognjena prevajalna funkcija.

7.3 Najožji prenosni kanal

Nyquistov kriterij pravi, da je prenos brez intersimbolne interference takrat in le takrat, kadar je prekrita oziroma prepognjena prevajalna funkcija ekvivalentnega prenosnega kanala konstantna. Ta pogoj izpolnjuje veliko število prevajalnih funkcij. Postavi pa se vprašanje katera od teh ima najmanjšo pasovno širino, to pomeni, da rabimo za prenos najožji frekvenčni pas.



Slika 7.7 – Primera prevajalnih funkcij ekvivalentnega prenosnega kanala. Funkcija na levi ne izpolnjuje Nyquistovega kriterija, saj prekrita prevajalna funkcija ni konstantna. Prevajalna funkcija na desni izpolnjuje Nyquistov kriterij in omogoča prenos brez intersimbolne interference.

Iz slike 7.8 ni težko ugotoviti, da bo prepognjena prevajalna funkcija zagotovo konstantna, kadar je prevajalna funkcija ekvivalentnega prenosnega kanala konstantna na intervalu med $-f_s/2$ in $f_s/2$, zunaj njega pa enaka 0, kar je karakteristika idealnega nizkega sita. V tem primeru namreč sploh ne pride do pregibanja, oziroma se premaknjene prevajalne funkcije ne prekrivajo. Pasovna širina takega sita je enaka:

$$B = \frac{f_s}{2} \quad (7.10)$$

Preprosto ugotovimo tudi, da je to najmanjša pasovna širina, ki še dopušča prenos brez intersimbolne interference. Očitno je namreč, da prekrita prevajalna funkcija ne more biti konstantna, če pasovno širino zmanjšamo pod $f_s/2$. Iz zgornje enačbe potem tudi preprosto ugotovimo, da je maksimalna simbolna hitrost preko kanala s pasovno širino B enaka dvakratni pasovni širini:

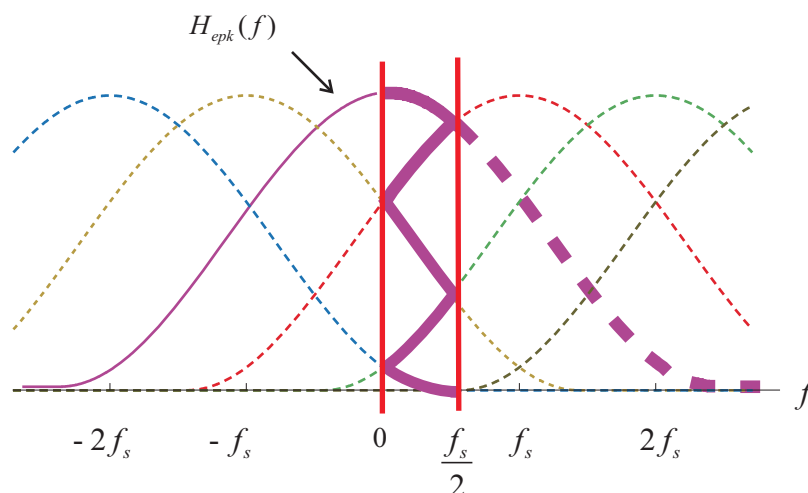
$$f_{smax} = 2B \quad (7.11)$$

To pomeni, da je mogoče brez intersimbolne interference prenašati največ 2 simbola na sekundo na vsak Hz pasovne širine, ki jo imamo na volj za prenos.

Teoretično lahko prenašamo brez intersimbolne interference 2 simbola na sekundo preko vsakega Hz pasovne širine, ki jo imamo na volj za prenos. Vendar je to zgolj teoretična zgornja meja, ki se ji lahko približamo, ne moremo pa je doseči, saj bi za to potrebovali idealno nizko sito, ki pa v praksi ni izvedljivo. Pasovna širina, ki jo potrebujemo za prenos je zato vedno nekoliko večja od teoretične meje.

7.4 Očesni diagram

Ali prihaja pri prenosu do intersimbolne interference lahko ugotovimo s pomočjo osciloscopa. Preko prenosnega kanala prenašamo naključen niz simbolov $y[n]$. To so lahko binarni ali pa tudi



Slika 7.8 – S prekrivanjem premaknjenih prevajalnih funkcij se prevajalna funkcija navidezno prepogiba okrog frekvence $f_s/2$. To imenujemo tudi prepognjena prevajalna funkcija. Ker je prekrita prevajalna funkcija periodična, zadošča, da preverimo konstantnost prepognjene prevajalne funkcije.

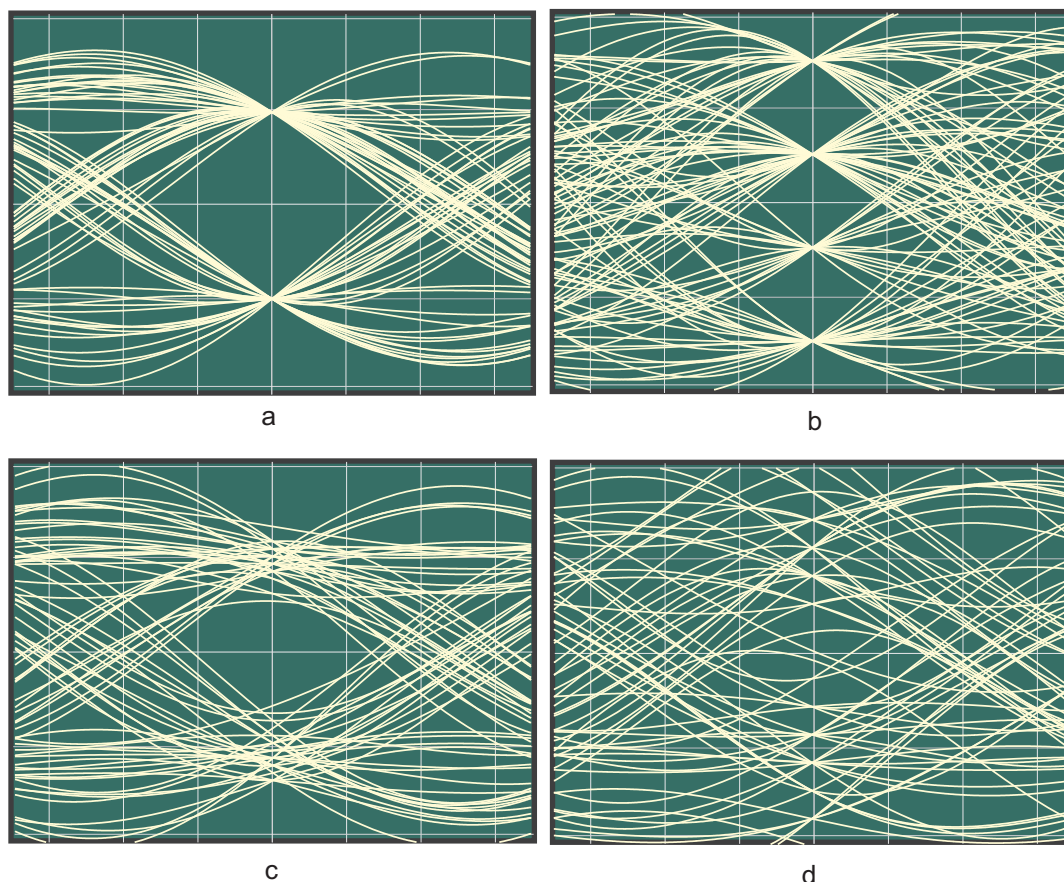
večnivojski simboli. Na vhod navpične odklonske osi ali tudi y osi osciloscopa vodimo signal $\hat{y}(t)$ iz izhoda prenosnega prenosnega kanala. Časovno skalo nastavimo tako, da traja en prelet žarka med 1,5 in 2 simbolna intervala T_s . Časovno bazo prožimo z zunajjim signalom, ki je sinhroniziran z digitalnim taktom oziroma simbolno frekvenco f_s . Ker je vhodni niz $y[n]$ naključen, so prehodi med posameznimi nivoji tudi naključni. Ker ima zaslon osciloscopa persistenco lahko naenkrat opazujemo večje število preletov žarka, ki ustrezajo različnim zaporedjem oddanih simbolov. V ta namen lahko uporabimo tudi spominski osciloscop, ki na zaslonu ohranja vse pretekle prelete, dokler zaslona ne osvežimo.

Nekaj primerov zaslonskih slik, ki jih lahko dobimo, je prikazanih na sliki 7.9. Kadar ni intersimbolne interference nastopajo ob trenutkih, ko bi signal vzorčili, samo točno določene vrednosti, ki ustrezajo vrednostim oddanih simbolov. Na zaslonu dobimo sliko, ki spominja na oko, zato imenujemo to tudi očesni diagram. Na sliki 7.9 a je prikazan očesni diagram pri prenosu binarnega niza brez intersimbolne interference, na sliki 7.9 b pa očesni diagram pri prenosu štirinivojskih simbolov brez intersimbolne interference. V obeh primerih pravimo da so očesa popolnoma odprta.

Če pride pri prenosu do intersimbolne interference se začne oko zapirati. Navpična razdalja med posameznimi nivoji se zmanjšuje. Na sliki 7.9 c je prikazan očesni diagram pri relativno majhni intersimbolni interferenci, ki sama po sebi še ne povzroča napak, saj je oko še vedno delno odprto. Je pa tak prenos bolj občutljiv na šum, saj šum dodatno zapira oko.

Če je intersimbolna interferenca še večja, se oko popolnoma zapre (slika 7.9 d) in ni več mogoče ločiti posameznih nivojev. V tem primeru prihaja pri prenosu nujno do napak.

Poleg višine odprtine očesa je pomembna tudi njena širina. Od širine je namreč odvisno, kako natančno je potrebno določiti vzorčne trenutke. Če je oko ožje morajo biti vzorčni trenutki bolj natančni, prenos pa je bolj občutljiv na tresenje faza, do katerega pride pri sinhronizaciji digitalnega takta.



Slika 7.9 – Primeri očesnih diagramov. a – binarno oko brez intersimbolne interference, b – štirinivojsko oko brez intersimbolne interference, c – zaradi intersimbolne interference delno zaprto binarno oko, oba nivoja se še ločita in so odločitve lahko pravilne, d – popolnoma zaprto binarno oko, pri odločitvah pride nujno do napak

7.5 Optimalno sprejemo sito

Doslej smo obravnavali zgolj brezšumen prenos, v praksi pa se koristnemu signalu pri prenosu doda šum, to je naključen signal, ki nastane kot posledica termičnega gibanja elektronov (termični šum) ali pa so to signali drugih, motilnih izvorov. Šum se običajno prišteva koristnemu signalu in govorimo o aditivnem šumu. Šum lahko povzroča napake pri prenosu. Koliko bo teh napak je odvisno od razmerja med močjo signala in močjo šuma (SNR – Signal to Noise Ratio), in vrste šuma, o čemer bo več govora v naslednjem poglavju. Zaenkrat naj zadošča, da želimo imeti pri sprejemu čim večje razmerje med signalom in šumom.

V modelu za prenos impulzov v osnovnem pasu na sliki 7.4 nastopajo v ekvivalentnem prenosnem kanalu tri sita. Z oblikovanjem impulzov in oddajnim sitom vplivamo na spekter oddajnega signala. Večinoma moramo poskrbeti, da prenos poteka v predpisanem frekvenčnem področju in, da je gostota spektra izven tega področja pod določeno predpisano vrednostjo. Kadar imamo na voljo svojo zasebno prenosno pot (najeta linija), ta vidik ni tako pomemben.

Sprejemno sito pa lahko načrtujemo tako, da prepušča čim več signala in čim manj šuma in na ta način maksimiramo razmerje med signalom in šumom. Sprejemno sito, ki maksimira

razmerje med signalom in šumom imenujemo optimalno sprejemno sito.

Šum, ki se prišteva signalu na prenosnem kanalu ima običajno zelo širok in raven spekter – to je tako imenovan bel šum. Ker je šum neodvisen od signala je korelacija šuma in signala majhna. Prisotnost nekega znanega signala v šumu zato najlažje ugotovimo s korelacijo. Z njo primerjamo šumen signal z znanim signalom. Če v šumu ni znanega signala je korelacija majhna, če pa je znan signal prisoten, je korelacija velika. Če ne vemo točno, v katerem času nastopa signal, moramo delati korelacijsko funkcijo. Omenimo še, da je sprejem s korelacijo optimalen. Dokaz za to je nekoliko zahtevnejši in ga bomo izpustili.

Predpostavimo sedaj, da smo oddali en sam simbol v obliki enotnega impulza. Zaradi preprostejših izrazov označimo sistemsko funkcijo ekvivalentnega prenosnega kanala brez sprejemnega sita s $h_k(t)$. Velja:

$$h_k(t) = g(t) * h_{odd}(t) * h_{pp}(t) \quad (7.12)$$

Ker je na vhodu tega kanala enotin impulz je signal na vhodu sprejemnega sita enak odzivu tega dela kanala na enotin impulz, to je njegovi sistemski funkciji $h_k(t)$. Temu signalu pa bo prištet še šum. Da bi najboljše detektirali prisotnost signala, želimo na sprejemni strani narediti korelacijsko funkcijo s sprejetega signala s $h_k(t)$.

Spomnimo se sedaj, da je signal na izhodu sita (linearnega časovno nespremenjivega sistema) enak konvoluciji signala in sistemske funkcije sita. Spomnimo se tudi, da je konvolucija enaka korelacijski funkciji s časovno obrnjenim signalom, oziroma obratno, korelacijo lahko izračunamo kot konvolucijo s časovno obrnjenim signalom:

$$r_{xy}(t) = x(t) * y(-t) \quad (7.13)$$

V našem primeru to pomeni, da moramo na sprejemu narediti konvolucijo s po času obrnjeno sistemsko funkcijo kanala, to je s $h_k(-t)$, oziroma, da mora biti sistemska funkcija optimalnega sprejemnega sita enaka po času obrnjeni sistemski funkciji kanala do tja:

$$h_{spr\ opt}(t) = h_k(-t) \quad (7.14)$$

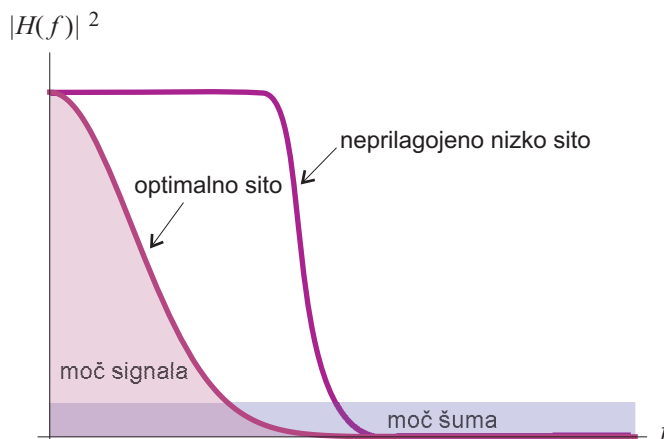
Naredimo še Fourierov transform zgornjega izraza in dobimo:

$$H_{spr\ opt}(f) = H_k^*(f) \quad (7.15)$$

Prevajalna funkcija optimalnega sprejemnega sita je enaka konjugirano kompleksni prevajalni funkciji celotnega kanala do sprejemnega sita.

Rezultat je morda na prvi pogled presenetljiv, saj pravi, da je amplitudna karakteristika sprejemnega sita kar enaka amplitudni karakteristiki kanala, oziroma, da signal še dodatno oslabimo točno tam, kjer je bil oslavljen že na kanalu. Vendar postane stvar bolj razumljiva, če jo pogledamo z drugega zornega kota. Vzemimo, da je karakteristika kanala nizkopasovna, šum pa kot smo že omenili, je širokopasoven, kot je to prikazano na sliki 7.10.

Vsako nizko sito, ki oslabi šum izven frekvenčnega pasu signala, izboljša razmerje med signalom in šumom. Optimalno sito pa je prilagojeno spektru signala. V področju, kjer je razmerje med signalom in šumom veliko, prepušča bolj, kjer je majhno pa bolj slabi. Na ta način dosežemo še boljše razmerje med signalom in šumom na izhodu kot z neprilagojenim nizkim sitom.



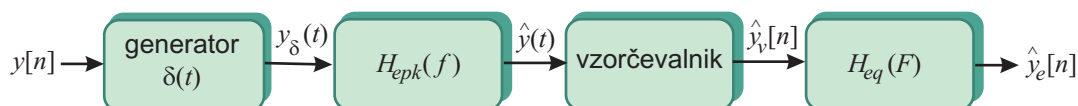
Slika 7.10 – Če neprilagojeno nizko sito izboljša razmerje med signalom in šumom, saj zaduši šum pri frekvencah, kjer ni signala. Optimalno sito pa je prilagojeno kanalu in s tem spektru sprejetega signala in zato daje še boljše razmerje med signalom in šumom.

7.6 Izravnava karakteristike kanala

Ko smo obravnavali optimalno sprejemno sito, smo predpostavili, da prenašamo en sam simbol. V resnici prenašamo niz simbolov in lahko pride do intersimbolne interference. Da do nje ne bi prišlo, mora biti prekrita prevajalna funkcija kanala konstantna, česar pa optimalno sprejemno sito ne zagotavlja. Zato je potrebno dodati še eno sito, ki bo izravnalo prekrito prevajalno funkcijo kanala. Tako sito imenujemo tudi izravnalnik (ang. equalizer). Prevajalna funkcija izravnalnika $H_{eq}(f)$ mor biti enaka inverzni prekrito prevajalni funkciji, da bi bil njun produkt konstanten. Ker je prekrito prevajalna funkcija po frekvenci periodična funkcija, mora biti tudi prevajalna funkcija izravnalnika periodična po frekvenci, zato jo izvedemo z digitalnim sitom po vzorčenju signala. Prevajalna funkcija digitalnega izravnalnika mora biti potem enaka:

$$H_{eq}(F) = \frac{AT_s}{\sum_{k=-\infty}^{\infty} H_{epk}(f_s(F+k))} \quad (7.16)$$

Kjer smo upoštevali da je F normirana frekvenca in velja $f = f_s F$. Vključitev izravnalnika v sprejemnik je prikazana na sliki 7.11. Čeprav je v vzorcih signala $y_v[n]$ še prisotna intersimbolna



Slika 7.11 – Izravnalnik prekrite prevajalne funkcije ima periodično prevajalno funkcijo, zato ga vključimo kot digitalno sito za vzorčevalnikom.

interferenca je v simbolih na izhodu izravnalnika $y_e[n]$ ni več, ker je izravnalnik izravnal prekrito prevajalno funkcijo kanala.

Če bi v času načrtovanja prenosnega sistema poznali prevajalno funkcijo kanala, bi lahko izračunali njegovo prevajalno funkcijo in načrtovali digitalno sito z ustrežno prevajalno funkcijo.

Običajno pa ob času načrtovanja ne poznamo karakteristike kanala, saj ne vemo točno na kakšen kanalu bo potekal prenos. Enak prenosni sistem lahko priključimo na kanale z zelo različnimi karakteristikami. Karakteristika kovinskega vodnika je na primer odvisna od njegove dolžine, premera vodnika, razdalje med vodniki, izolacije in podobno. Poleg tega se spreminja tudi s temperaturo, kar pomeni, da se lahko spreminja tudi med samim delovanjem sistema.

Da bi prišli do referenčnega signala moramo vključiti v prenosni sistem še odločitveno vezje, ki smo ga omenili že v modelu digitalnega prenosa, ki je prikazan na sliki 7.1.

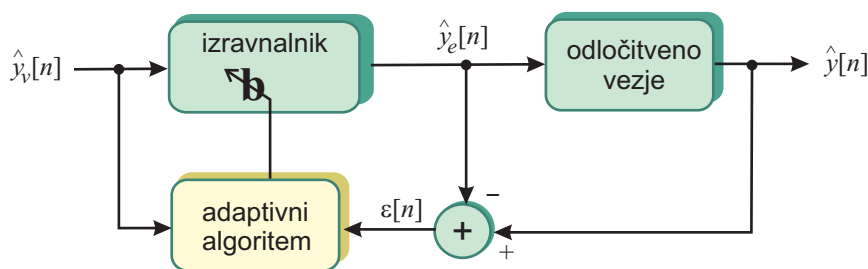
7.7 Adaptiven izravnalnik

Ker običajno vnaprej ne poznamo karakteristike kanala, preko katerega bomo prenašali, moramo izvesti izravnalnik adaptivno. Izvedemo ga lahko z adaptivnim FIR sitom, ki smo ga obravnavali v razdelku 6.2. Pri adaptivnem situ potrebujemo referenčni signal $d[n]$, da lahko potem z ustreznim algoritmom prilagajamo uteži tako, da je izhodni signal čim bolj podoben referenčnemu signalu oziroma da minimiziramo srednjo kvadratično vrednost napake.

Pri izravnalniku na sliki 7.11 smo signal na izhodu izravnalnika označili z $\hat{y}_e[n]$. Želimo, da je ta signal enak signalu $y[n]$, to je signalu ki ga prenašamo. Signal $y[n]$ je torej referenčni signal, vendar ga pri sprejemu nimamo na voljo. Če bi ga imeli na voljo, potem ga pač ne bi bilo potrebno prenašati.

Da bi lahko pridobili referenčni signal, moramo v sistem vključiti še odločitveno vezje, ki smo ga že omenili pri obravnavi modela digitalnega prenosa, ki je prikazan na sliki 7.4. Odločitveno vezje se na osnovi signala na svoje vходу odloča, kateri simboli so bili oddani, tako, da je, če so te odločitve pravilne, signal na njegovem izhodu $\hat{y}[n]$ enak oddanemu signalu $y[n]$ in lahko služi kot referenčni signal.

Vključitev adaptivnega izravnalnika na osnovi FIR sita na sprejemni stani je prikazana na sliki 7.12. Prikazan je samo del sistem za vzorčevalnikom na sliki 7.11.



Slika 7.12 – Adaptiven izravnalnik na osnovi FIR sita. Kot referenčni signal služi signal $\hat{y}[n]$ na izhodu odločitvenega vezja, ki je približek pravemu referenčnemu signalu $y[n]$, to je signalu, ki ga prenašamo. Če je več odločitev odločitvenega vezja pravilnih kot napačnih, se uteži \mathbf{b} FIR sita popravljajo v pravi smeri.

Pri nekodiranem prenosu je odločitveno vezje preprost večnivojski komparator, ki sprejeti vzorec $\hat{y}_e[n]$ primerja z vrednosti, ki pripadajo abecedi prenašanih simbolov, in se odloči, da je sprejet tisti simbol iz abecede, ki je po vrednosti najbližje sprejetemu vzorcu. Pri dvonivojskem prenosu $y[n] \in \{-1, 1\}$ se na primer odloči, da je sprejeti simbol 1, kadar je vzorec večji od

nič in -1, kadar je manjši od nič. Kadar imamo kodiran prenos, vsebuje odločitveno vezje tudi dekodirnik. O kodiranju in dekodiranju bo več govora v naslednjih poglavjih.

Če intersimbolna interferenca in šum na začetku nista prevelika, potem je večina odločitev pravilnih že na začetku, tako da se uteži FIR sita \mathbf{b} popravljajo proti optimalni vrednosti. S tem se zmanjšuje intersimbolna interferenca, kar še izboljša naslednje odločitve.

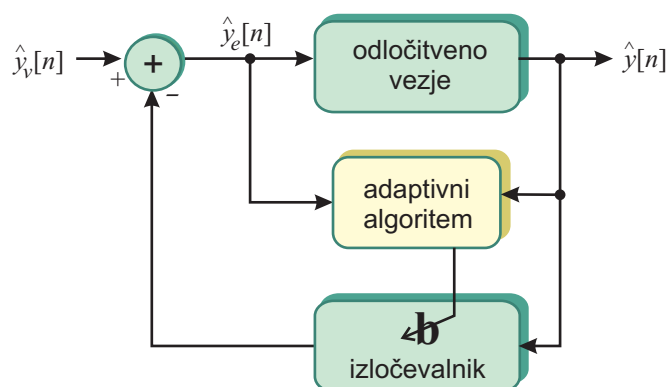
Da bi izločil intersimbolno interferenco, adaptivni izravnalnik izravnava prekrito prevajalno funkcijo ekvivalentnega prenosnega kanal. Ker smo v to funkcijo pred tem že vključili optimalno sprejemno sito, ki maksimira razmerje med signalom in šumom SNR, izravnalnik to razmerje nekoliko pokvari.

Adaptivni algoritem prilagaja uteži sita tako, da je srednja kvadratična vrednost napake čim manjša. Napako povzročata tako intersimbolna interferenca kot tudi šum. Sito se v stacionarnem stanju nastavi tako, da je njuna skupna moč najmanjša, kar je lahko celo nekoliko boljše kot samo izravnava prekrite prevajalne funkcije.

7.8 Adaptiven izločevalnik intersimbolne interference

Pri sprejemu se lahko odločimo, da bomo, namesto, da bi izravnali prekrito prevajalno funkcijo kanala, intersimbolno interferenco raje odšteli od sprejetega signala. Ker pri tem ne spreminjamo prevajalne funkcije, tudi ne pokvarimo razmerja med signalom in šumom.

Izločevalnik intersimbolne interference sloni na predpostavki, da se kanal spreminja relativno počasi in, da zato isto zaporedje oddanih simbolov v nekem krajšem časovnem obdobju povzroči vedno isto intersimbolno interferenco. Če torej poznamo simbole, ki so bili oddani pred trenutnim simbolom, lahko iz njih izračunamo intersimbolno interferenco, ki so jo povzročili, in jo odštujemo od sprejetega signala, kot je to prikazano na sliki 7.13.



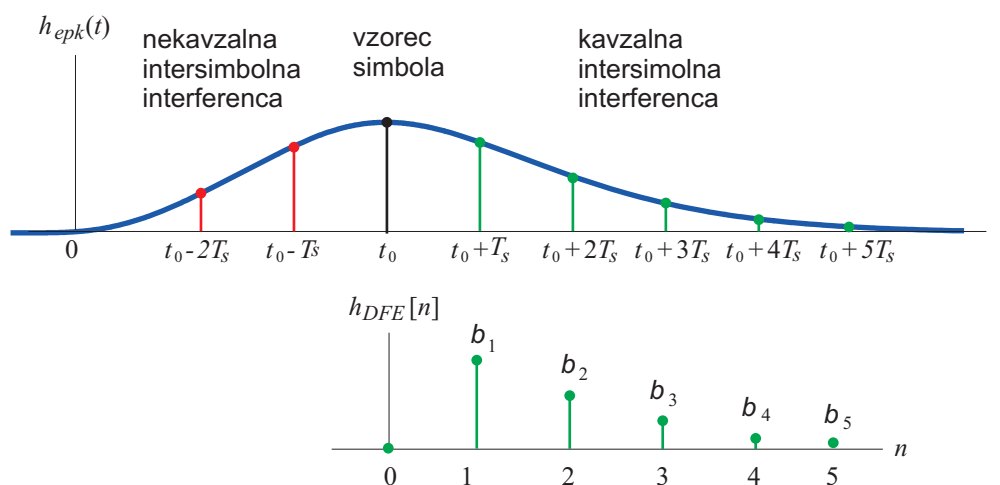
Slika 7.13 – Adaptiven izločevalnik intersimbolne interference. Na vhod adaptivnega sita vodimo odločitve o preteklih sprejetih simbolih $\hat{y}[n]$. Če so te odločitve pravilne, bomo pravilno izračunali intersimbolno interferenco, ki jo povzročijo, in jo odšteli od sprejetega signala $\hat{y}_v[n]$

Na vhod adaptivnega sita vodimo odločitve o $\hat{y}[n]$. Uteži adaptivnega sita prilagajamo na enak način kot pri izravnalniku, tako da zmanjšujemo razliko med sprejetim signalom $\hat{y}_v[n]$ in referenčnim signalom $\hat{y}[n]$. Ta razlika bo najmanjša, kadar od sprejetega signala odštujemo vso

intersimbolno interferenco, ki jo povzročajo simboli na vhodu adaptivnega sita. Ker odločitve o sprejetem signalu vodimo nazaj na vhod, imenujemo izločevalnik intersimbolne interference tudi izravnalnik z odločitvijo v povratni vezavi (DFE – Decision Feedback Equalizer).

Ker so odločitve o sprejetem signalu tudi na vhodu FIR sita, povzročijo napačne odločitve bistveno večjo napako, kot pri izravnalniku, saj odštejemo napačno izračunano vrednost intersimbolne interference in jo lahko na ta način še povečamo, kar lahko povzroči nove napačne odločitve. Ta pojav imenujemo širjenje napake. Da ne bi prišlo do širjenja napake, morajo biti pri izločevalniku intersimbolne interference razmere na začetku bistveno boljše, kot je to potrebno pri izravnalniku.

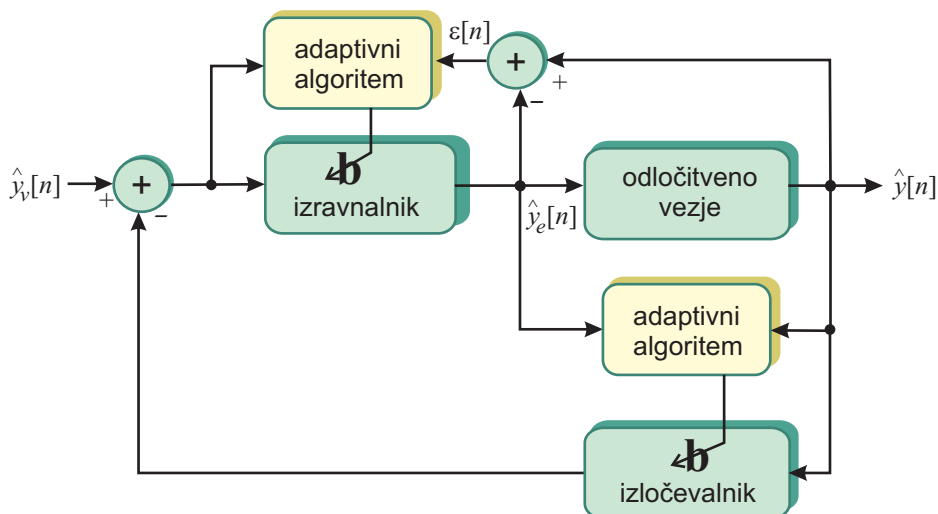
Druga pomanjkljivost izločevalnika intersimbolne interference je, da je z njim mogoče izločiti samo intersimbolno interferenco, ki so jo povzročili pretekli simboli, o katerih smo se že odločili (kavzalna intersimbolna interferenca, ne more pa izločiti intersimbolne interference prihodnjih simbolov (nekavzalna intersimbolna interferenca). Zaradi zakasnitve na kanalu lahko prihodnji simboli ravno tako povzročajo intersimbolno interferenco, kot je to prikazano na slik 7.14.



Slika 7.14 – Ker čas vzorčenja prilagodimo zakasnitvi na prenosni poti tako, da vzorčimo pri maksimalni vrednosti odziva, lahko intersimbolno interferenco povzročajo tudi simboli, ki jih še nismo vzorčili. To imenujemo nekavzalna intersimbolna interferenca. Da bi izločili kavzalno intersimbolno interferenco, morajo biti sistemska funkcija FIR sita $h_{DFE}[n]$ enaka vzorcem kavzalne intersimbolne interference.

V stacionarnem stanju, ko je FIR sito prilagojeno, je sistemska njegova funkcija $h_{DFE}[n]$ enaka vzorcem kavzalne intersimbolne interference. Ker so pri FIR situ uteži b_n enake sistemski funkciji, pomeni, da se uteži nastavijo na vrednosti vzorcev kavzalne intersimbolne interference, oziroma vrednosti $h_{epk}(t)$ ob kavzalnih vzorčnih trenutkih.

Ker izločevalnik intersimbolne interference ni sposoben izločiti nekavzalne intersimbolne interference se v praksi pogosto uporablja kombinacija izravnalnika in izločevalnika, kot je to prikazano na sliki 7.15.



Slika 7.15 – Kombinacija izravnalnika in izločevalnika intersimbolne interference.

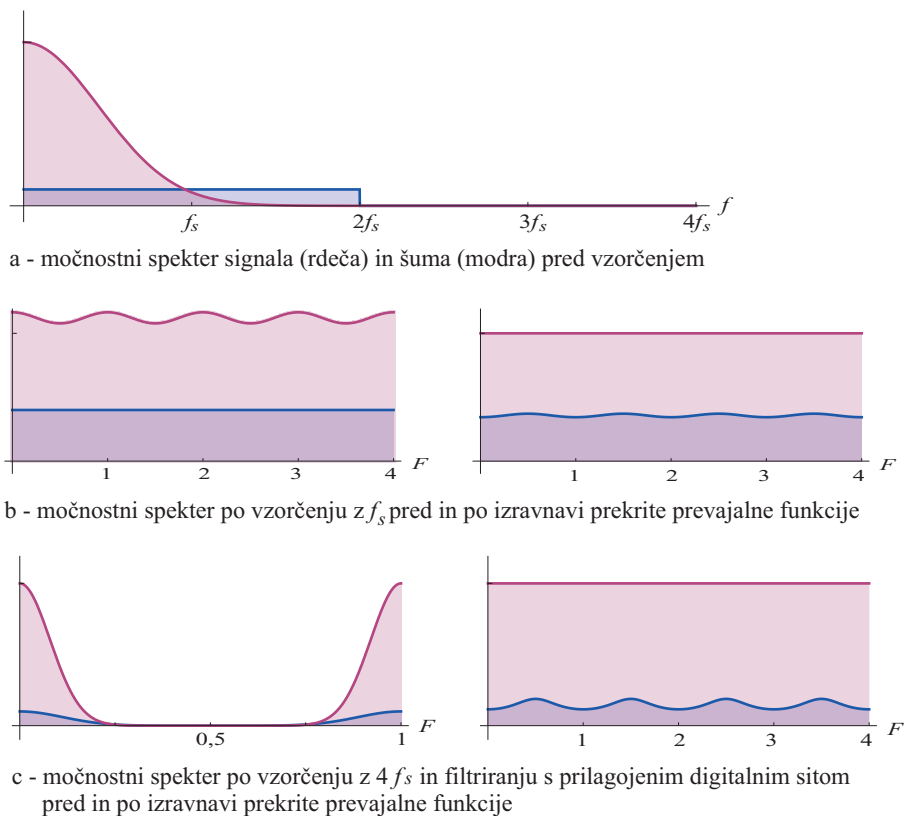
7.9 Izravnalnik s prevzorčenjem

Na sprejemni strani smo vključili optimalno sprejemno sito ki maksimira razmerje med signalom in šumom. Njegova prevajalna funkcija mora biti v skladu z izrazom (7.15) enaka konjugirano kompleksni prevajalni funkciji ekvivalentnega prenosnega kanala brez sprejemnega sita $H_k^*(f)$. To je aperiodična funkcija, ki šum v frekvenčnem področju, kjer je moč signala nizka. Po vzorčenju s simbolno frekvenco se področja, kjer ni signala in področja signal prekrijejo in šuma ni več mogoče ločiti od signala. Zato moramo optimalno sito izvesti v analogni obliki in ga vključiti pred vzorčevalnik. Analogno sito pa ne more biti adaptivno in ga ne moremo prilagoditi karakteristiki prenosne poti, če te vnaprej ne poznamo.

Želeli bi, da je tudi to sito adaptivno, da bi se lahko prilagodilo kanalu, na katerega ga priključimo, oziroma da bi se lahko sproti prilagajalo spremembam kanala. Zato mora biti to sito digitalno in vključeno za vzorčevalnik. Da pa se po vzorčenju šum in signal ne prekrijeta popolnoma, moramo povečati vzorčno frekvenco. Pred vzorčenjem moramo signal omejiti z nizkim sitom z mejno frekvenco manjšo ali enako $f_{vz}/2$, da izločimo šum, ki bi se po vzorčenju preslikal v frekvenčni pas signala. Na ta način lahko po vzorčenju s prilagojenim sitom izločijo šum izven frekvenčnega pasu kanala s tako imenovanim izravnalnikom s prevzorčenjem (ang. fractionally spaced equalizer).

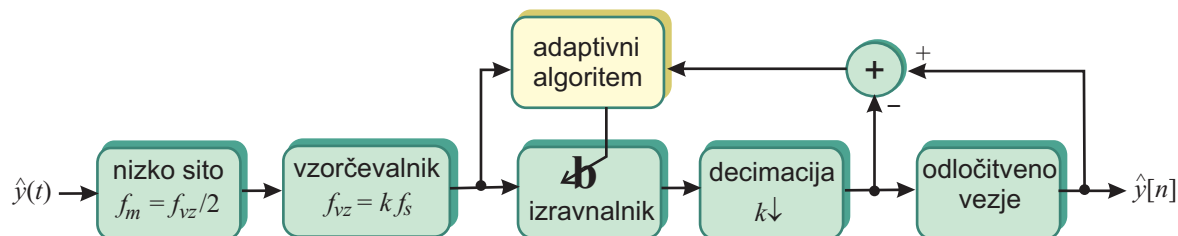
Razmere so prikazane na sliki 7.16, kjer je rdeče prikazan močnostni spektre signala, modro pa močnostni spekter šuma. Ploščina pod med absciso in krivuljo močnostnega spektra je po Parsevalovem teoremu enaka moči signala. Slika 7.16.a – Signal in šum sta pred vzorčenjem omejena na pasovno širino $B = 2f_s$. Slika 7.16.b – Pri vzorčenju s simbolno frekvenco f_s se signal in šum prekrijeta, po izravnavi kanala pa se šum še nekoliko ojači. Slika 7.16.c – Če je je vzorčna frekvenca štirikrat višja ($f_{vz} = 4f_s$), lahko s prilagojenim sitom po vzorčenju izboljšamo razmerje med signalom in šumom. Nato vzamemo vsak četrti vzorec za detekcijo simbolov, kar ustreza vzorčenju s simbolno frekvenco f_s . Prevajalna funkcija se ponovno prekrije, po izravnavi

se razmerje med signalom in šumom nekoliko poslabša, vendar, če primerjamo ploščine pod krivuljo šuma na desni strani slik c in d, vidimo, da je to v drugem primeru bistveno višje. Z dovolj visoko vzorčno frekvenco, pri kateri zajamemo večino moči signala, se lahko popolnoma približamo optimalnemu situ. Vezava izravnalnika s prevzorčenjem je prikazana na sliki 7.17.



Slika 7.16 – Močnostni spekter pri vzorčenju s simbolno hitrostjo in pri prevzorčenju in filtriranju s prilagojenim sitom.

Vzorčno frekvenco povečamo za faktor k , tako, da velja $f_{vz} = kf_s$. Pred vzorčevalnik vključimo



Slika 7.17 – Izravnalnik s prevzorčenjem opravlja hkrati funkcijo optimalnega sita in izravnalnika prekrite prevajalne funkcije.

nizko sito z mejno frekvenco $f_m = f_{vz}/2$, da izločimo šum nad to frekvenco, ki bi se preslikal nazaj v osnovni pas. Obe siti, to je prilagojeno sito in sito za izravnavo prekrite prevajalne funkcije izvedemo z enim adaptivnim FIR sitom. Adaptivni algoritem namreč minimizira srednjo kvadratično vrednost napake, ki je posledica tako šuma, kot intersimbolne interference, zato se

bodo uteži nastavile tako, da bodo maksimirale razmerje med signalom in šumom. To pa je natanko to kar želimo. Ker imamo na vsak sprejeti simbol k vzorcev, vključimo pred odločitveno vezje decimacijo za faktor k , to pomeni, da med dvema vzorcema izpustimo $k - 1$ vzorcev.

7.10 Spekter signala v osnovnem pasu

Primere različnih signalov pri digitalnem prenosu v osnovnem pasu smo si ogledali na začetku poglavja in so prikazani na sliki 7.2. Kakšen je signal je, ki ga dobimo iz signalnega generatorja je močno odvisno od izbrane oblike impulzov. Predpostavimo, da so simboli $y[n]$ na vhodu impulznega generatorja med seboj neodvisni oziroma da velja:

$$R_{yy}[m] = \overline{y^2[m]} \delta[m] \quad (7.17)$$

kar pomeni, da je avtokorelacijska funkcija enaka 0 pri vsakem zamiku signala $m \neq 0$. Močnostni spekter je enak TDFT avtokorelacijske funkcije in je torej enak:

$$S_y(F) = \overline{y^2[m]} \quad (7.18)$$

Spekter je konstanten, neodvisen od frekvence F , oziroma bel. Kasneje bomo videli, da lahko spekter digitalnega signala preoblikujemo z ustreznim kodiranjem.

Generator impulzov smo v razdelku 7.2 razdelili na dva dela, na generator enotinih impulzov in sito za oblikovanje impulzov. Močnostni spekter signala na izhodu generatorja enotinih impulzov je potem v skladu z izrazom (5.11) tudi bel in enak:

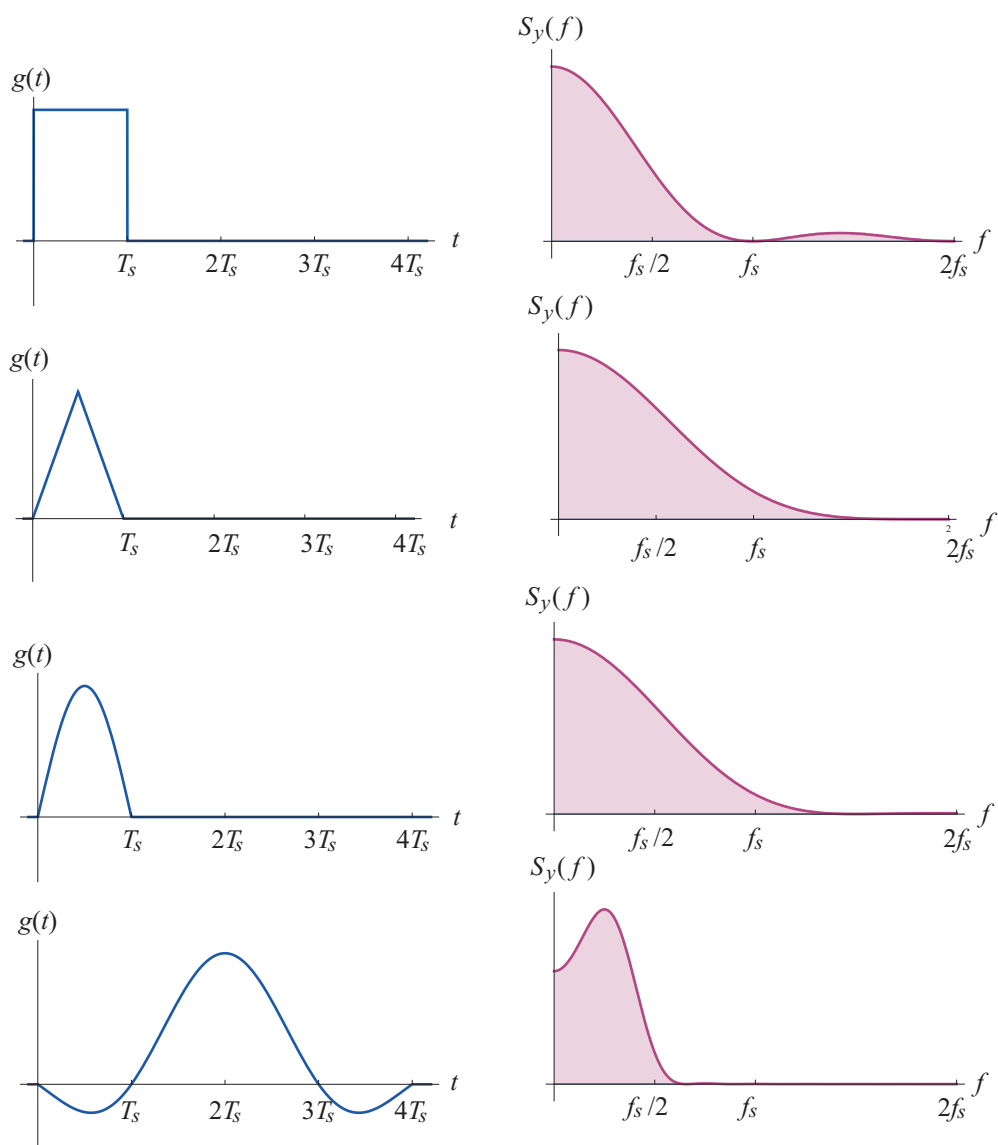
$$S_{y_s}(f) = \overline{y^2[m]} \quad (7.19)$$

Na prehodu skozi sito za oblikovanje s sistemsko funkcijo enako obliki impulza $g(t)$, se močnostni spekter signala množi s kvadratom absolutne vrednosti prevajalne funkcije $G(f)$:

$$S_y(f) = \overline{y^2[m]} |G(f)|^2 \quad (7.20)$$

kjer je $G(f)$ Fourierov transform $g(t)$. Močnostni spekter nekodiranega digitalnega signala pri različnih oblikah impulzov je prikazan na sliki 7.18.

Z oddajnim sitom lahko spekter še dodatno oblikujemo, običajno z njim porežemo spektralne komponente, ki so izven frekvenčnega pasu namenjenega za prenos, vendar s tem v splošnem vnašamo tudi intersimbolno interferenco.



Slika 7.18 – Močnostni spekter digitalnega signala pri različnih oblikah impulzov. Z obliko impulza močno vplivamo na spekter signala. Pravokotni impulzi, ki so sicer preprosti za izvedbo, imajo relativno zelo širok spekter.

Digitalni modulacijski postopki

Poglavje obravnava digitalne modulacijske postopke. Digitalni modulacijski postopki so načeloma enaki kot analogni modulacijski postopki, od njih se razlikujejo predvsem v modulacijskem signalu. Predstavljeni so postopki ASK, FSK, PSK, PAM, CPM, QAM, QPSK, MSK, GMSK in OFDM.

V prejšnjem poglavju smo obravnavali digitalni prenos v osnovnem pasu. Prenos v višjih frekvenčnih legah izvedemo tako, da kot modulacijski signal v amplitudni, frekvenčni ali fazni modulaciji uporabimo signal, ki ga dobimo iz generatorja impulzov pri prenosu v osnovnem pasu.

Kadar so impulzi pravokotni se amplituda, faza oziroma frekvenca moduliranega signala skokovito spreminjamo, zato govorimo o preklapljanju (ang. shift keying), modulacijski postopki pa se imenujejo amplitudno preklapljanje (ASK – Amplitude Shift Keying), frekvenčno preklapljanje (FSK – Frequency Shift Keying) ali fazno preklapljanje (PSK - Phase Shift Keying).

Kadar oblika impulzov ni pravokotna, se omenjene veličine spreminjajo zvezno s časom, kot pri impulzno amplitudni modulaciji (PAM - Pulse Amplitude Modulation), zvezni fazni modulaciji (CPM – Continuous Phase Modulation), Modulaciji z minimalnim faznim skokom (MSK – Minimum Shift Keying) in Gaussovi MSK (GMSK). Spektralno najučinkovitejši modulacijski postopek je kvadraturna modulacija (QAM - Quadrature Amplitude Modulation) medtem ko kapaciteto prenosne poti najbolj izkorišča prenos z več nosilci. Predstavnik takega prenosa sta ortogonalni frekvenčni multipleks (OFDM –Orthogonal Frequency Division Multiplex) in diskretni prenos z več nosilci (DMT - Discrete Multi Tone).

Pred kratice, ki označujejo posamezen modulacijski postopek pogosto pišemo številko, ki pove, koliko različnih vrednosti ima lahko posamezen prenašan simbol, kot npr 8PSK, 64QAM in podobno.

Poleg omenjenih seveda obstaja še cela vrsta drugih nekoliko modificiranih modulacijskih postopkov, vendar ti temeljijo na istih osnovah kot zgoraj omenjeni postopki.

Signal z modulacijskim postopkom, podobno kot pri analognem prenosu, najprej frekvenčno prestavimo na vmesno frekvenco f_{vm} in nato z mešanjem prestavimo v končno frekvenčno lego. Ker se mešanje pri digitalnem prenosu v ničemer ne razlikuje od mešanja pri analognem prenosu, ga tu ne bomo posebej obravnavali.

8.1 Amplitudna modulacija

Pri impulzno amplitudni modulaciji z nizom impulzov v osnovnem pasu moduliramo amplitudo nosilca:

$$y(t) = 2x(t)\cos(2\pi f_{vm}t) \quad (8.1)$$

Vidimo, da je modulacijski postopek popolnoma ekvivalenten amplitudni modulaciji z zadušenim nosilcem (SCAM), le da tu modulacijski signal $x(t)$ ni analogen signal temveč je to niz impulzov, ki smo ga imeli pri prenosu v osnovnem pasu. Če pri prenosu ne želimo imeti odvečne moči v nosilcu, signal $x(t)$ ne sme imeti enosmerne komponente, oziroma mora veljati:

$$\overline{x(t)} = 0 \quad (8.2)$$

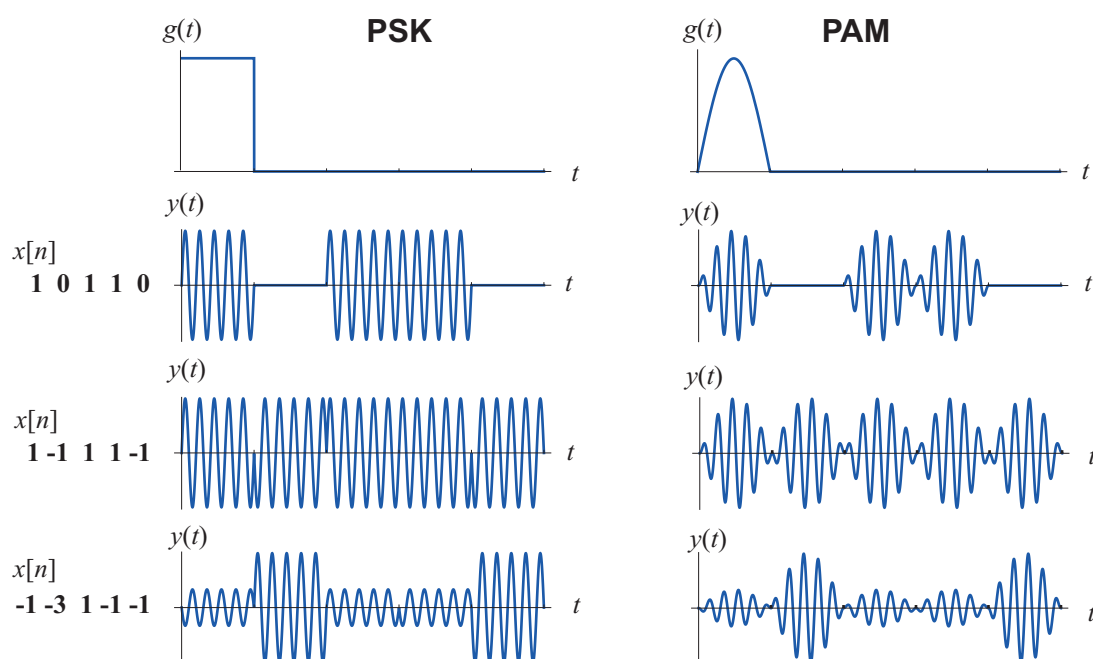
kar pa dosežemo če je digitalni signal $x[n]$ ki ga prenašamo enosmerno uravnotežen in velja:

$$\overline{x[n]} = 0 \quad (8.3)$$

Drug način za dosegost istega cilja, kadar $x[n]$ ni uravnotežen je, da uporabimo uravnotežene impulze $g(t)$ za katere velja:

$$\int_{-\infty}^{\infty} g(t) dt = 0 \quad (8.4)$$

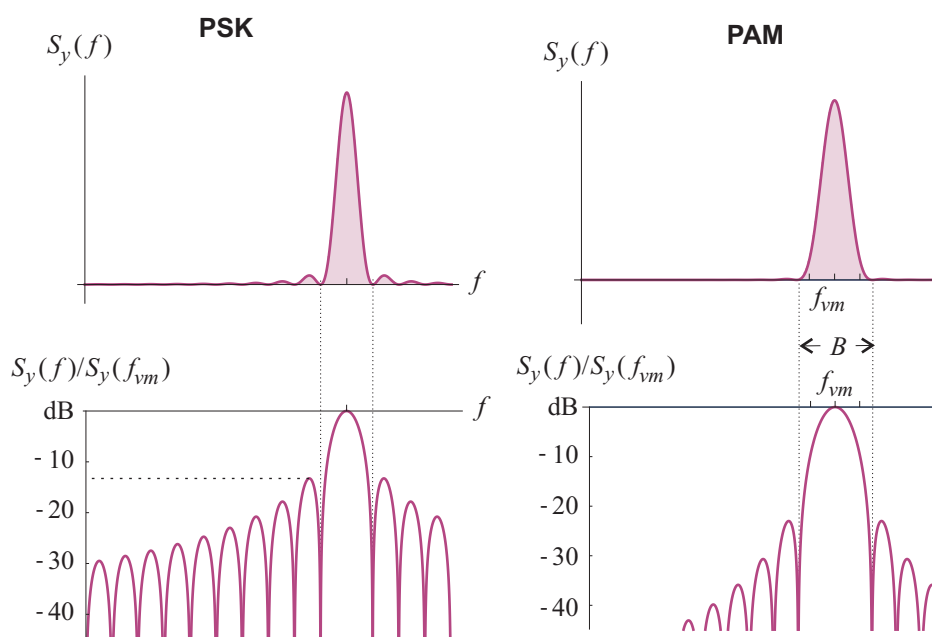
vendar imajo taki impulzi širši spekter. Modulacijo in demodulacijo izvedemo na popolnoma enak način kot pri SCAM. Na sliki 8.1 so prikazani modulirani signali pri različnih vhodnih abecedah in za pravokotno in pol sinusno obliko impulzov.



Slika 8.1 – Digitalno amplitudno modulirani signali pri pravokotnih impulzih (PSK) in pri pol-sinusnem impulzu (PAM) pri različnih abecedah digitalnega niza $\{0,1\}$, $\{-1,1\}$ in $\{-3,-1,1,3\}$

Kadar so impulzi pravokotne oblike imenujemo ta postopek amplitudno preklapanje (ASK) pri drugih oblikah impulzov pa impulzno amplitudna modulacija (PAM). Pri modulaciji se spekter iz osnovnega pasu preslika v levi in desni bok okrog nosilca. Pasovna širina se pri tem podvoji. Prenos je lahko dvonivojski ali več nivojski.

Močnostni spekter signala se pri modulaciji preslika levo in desno od nosilne frekvence. Pasovna širina signala se pri tem podvoji. Močnostni spekter PSK je prikazan na sliki 8.2. Osnovni snop spektra je pri ASK štirikrat širši od najožjega prenosnega kanala. Spekter je prikazan v linearnem in logaritemskem merilu - v dB. V logaritemskem merilu je namreč bistveno bolj viden del spektra izven osnovnega snopa, ki običajno predstavlja motnjo v sosednjih kanalih. Vidimo, da sta pri PSK prva stranska snopa oslabljena le za okrog 14 dB. Pri PAM s pol-sinusnimi impulzi je pasovna širina osnovnega snopa nekoliko večja, zato pa je sevanje izven nje bistveno manjše. Z impulzi, ki presegajo dolžino simbolnega intervala $T - S$ je možno doseči bistveno večje slabljenje izven pasu prenosa kakor tudi bistveno ožjo pasovno širino, kot smo to videli že na sliki 7.18.



Slika 8.2 – Močnostni spekter PSK in PAM s pol-sinusnimi impulzi. Pasovna širina se pri modulaciji podvoji. Sevanje izven osnovnega snopa pasovne širine $B = 2f_s$ je pri PSK dokaj veliko in lahko močno moti sosednje kanale. Zato PSK za prenos, kjer uporabljajo sosednje frekvence drugi uporabniki, ni najprimernejši modulaijski postopek. Pri PAM je osnovni snop nekoliko širši, zato pa je sevanje izven tega pasu precej manjše.

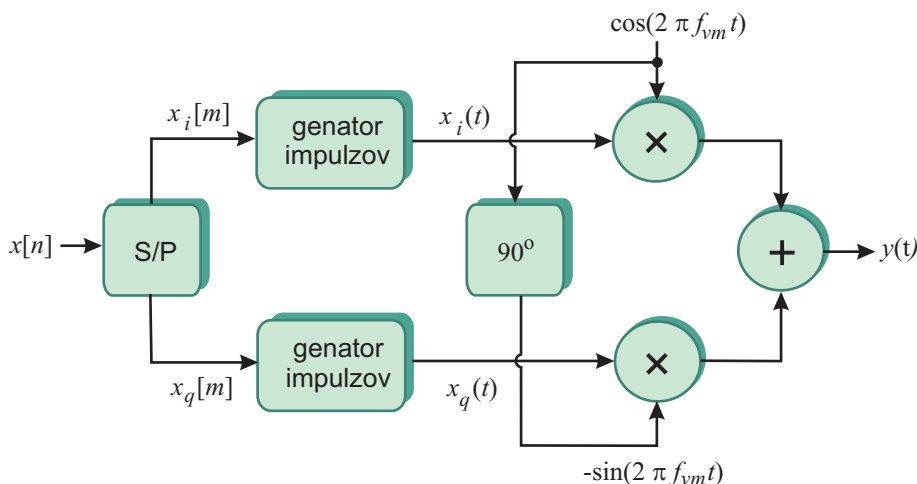
8.2 Kvadratura amplitudna modulacija

Pri kvadraturi amplitudni modulaciji (QAM – Quadrature Amplitude Modulation) moduli-ramo dva ortogonalna nosilca $\sin(2\pi f_{vm}t)$ in $\cos(2\pi f_{vm}t)$, vsakega s svojim signalom. Ker sta nosilca ortogonalna, ju je mogoče pri demodulaciji zopet ločiti. Moduliran signal pri kvadraturi

modulaciji lahko zapišemo v obliki:

$$y(t) = x_i(t)\cos(2\pi f_{vm}t) - x_q(t)\sin(2\pi f_{vm}t) \quad (8.5)$$

Kjer sta $x_i(t)$ in $x_q(t)$ dva niza impulzov v osnovnem pasu. Za prvega rečemo, da je v fazi (ang. in phase) in za drugega, da je v kvadraturi (ang. quadrature). Kvadrturni modulator je prikazan na sliki 8.3.



Slika 8.3 – QAM modulator. Vhodni niz simbolov $x[n]$ s simbolno hitrostjo f_s najprej na serijsko paralelnem pretvorniku razdelimo v dva niza simbolov $x_i[n]$ in $x_q[n]$ s polovično simbolno hitrostjo. Z generatorjema impulzov nato tvorimo signala $x_i(t)$ in $x_q(t)$, s katerima moduliramo ortogonalna nosilca. Izhodni signal dobimo kot vsoto signalov v obeh vejah.

Vhodni niz simbolov $x[n]$ s simbolno hitrostjo f_s najprej razdelimo v dva niza simbolov $x_i[n]$ in $x_q[n]$ s polovično simbolno hitrostjo $f_s/2$. Nato generiramo impulze, da dobimo signal v fazi $x_i(t)$ ki ga množimo z nosilcem (kosinus) in signal v kvadraturi $x_q(t)$, ki ga množimo z za 90° fazno zamaknjenim nosilcem (sinus). Signala v obeh vejah nato seštejemo, da dobimo QAM signal $y(t)$.

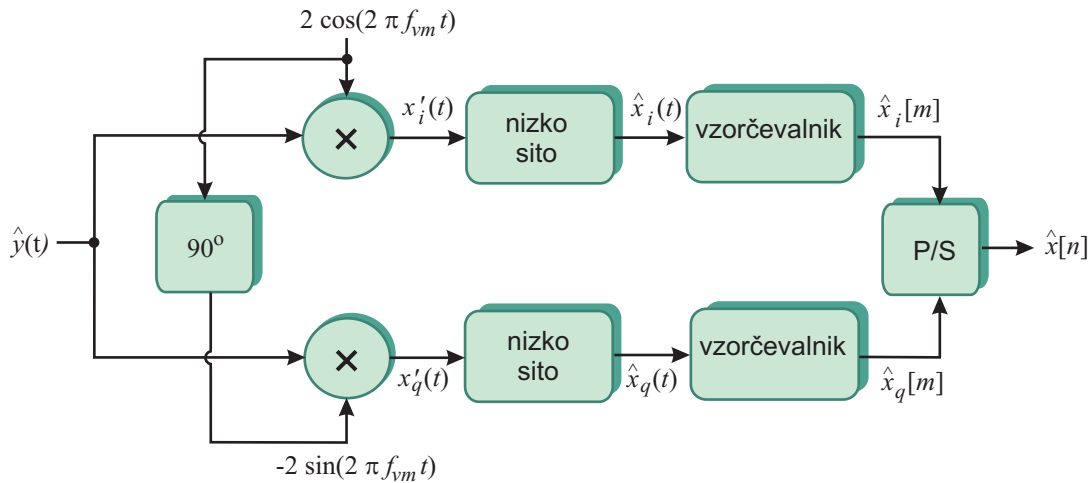
Ker se, ko razdelimo osnovni niz simbolov na dva niza, simbolna frekvenca razpolovi, je pasovna širina signalov $x_i(t)$ in $x_q(t)$ tudi za polovico manjša. Pri modulaciji (množenju s sinusom in kosinusom) se pasovna širina zopet podvoji. Končna pasovna širina je zato enaka, kot bi bila pri prenosu v osnovnem pasu in polovico manjša kot pri PAM.

Pri QAM potrebujemo za prenos enako pasovno širino kot pri prenosu v osnovnem pasu. QAM je spektralno najučinkovitejši modulacijski postopek

Demodulacijo izvedemo na enak način kot pri SCAM, le da demoduliramo vsako vejo posebej in potem oba niza vzorcev zopet združimo v en niz, kot je to prikazano na sliki 8.4.

Če je sprejeti signal $\hat{y}(t)$ enak oddanemu signalu $y(t)$, potem jetudi $\hat{x}[n]$ enak $x[n]$, saj velja:

$$x'_i(t) = 2y(t)\cos(2\pi f_{vm}t) = 2x_i(t)\cos^2(2\pi f_{vm}t) - 2x_q(t)\sin(2\pi f_{vm}t)\cos(2\pi f_{vm}t) =$$



Slika 8.4 – QAM demodulator. Sprejeti signal najprej razdelimo v dve veji in vsako vejo sinhrono demoduliramo. Nizko sito izloči komponente, ki nastopijo pri dvakratni nosilni frekvenci. Po vzorčenju oba niza zopet združimo s paralelnim serijskim pretvornikom.

$$\begin{aligned}
 &= x_i(t) + x_i(t) \cos(4\pi f_{vm}t) - x_q(t) \sin(4\pi f_{vm}t) \\
 x_i'(t) &= -2y(t) \sin(2\pi f_{vm}t) = -2x_i(t) \cos(2\pi f_{vm}t) \sin(2\pi f_{vm}t) + 2x_q(t) \sin^2(2\pi f_{vm}t) = \\
 &= x_q(t) - x_q(t) \cos(4\pi f_{vm}t) + x_i(t) \cos(4\pi f_{vm}t)
 \end{aligned} \tag{8.6}$$

Ker z nizkim sitom izločimo spekter pri dvojni frekvenci nosilca, ostane v obeh rezultatih le prvi člen. QAM signal v izrazu (8.5) lahko zapišemo tudi v obliki:

$$y(t) = A_y(t) \cos(2\pi f_{vm}t + \phi_y(t)) = \Re\{A_y(t) e^{j\phi_y(t)} e^{j2\pi f_{vm}t}\} \tag{8.7}$$

kjer sta $A_y(t)$ in $\phi_y(t)$ podana z:

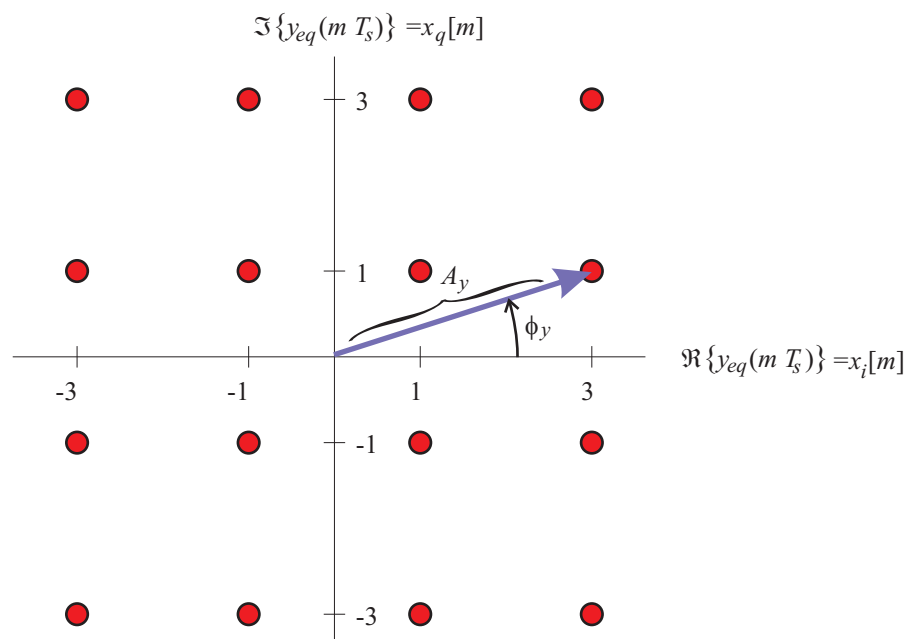
$$\begin{aligned}
 A_y(t) &= \sqrt{x_i^2(t) + x_q^2(t)} \\
 \phi_y(t) &= \arctan\left(\frac{x_q(t)}{x_i(t)}\right) + \frac{\pi}{2}(\text{sign}(x_i(t)) - 1)
 \end{aligned} \tag{8.8}$$

Kompleksen signal, ki nastopa pod znakom za realno vrednost v izrazu (8.7), je, kadar se $\phi_y(t)$ ne spreminja, kazalec, ki se vrti s frekvenco f_{vm} v kompleksni ravnini. Njegovo kompleksno amplitudo (to je amplitudo in fazo), ki se glede na frekvenco nosilca relativno počasi spreminjata, imenujemo tudi ekvivalentni signal v osnovnem pasu in ga označimo z $y_{eq}(t)$:

$$y_{eq}(t) = A_y(t) e^{j\phi_y(t)} \tag{8.9}$$

Iz zgornjih enačb je razvidno, da je QAM amplitudno fazna modulacija, saj se v odvisnosti od modulatorskih signalov $x_i(t)$ in $x_q(t)$ spreminjata tako amplituda $A_y(t)$ kot tudi faza $\phi_y(t)$ moduliranega signala.

Če ni intersimbolne interference lahko signala $x_i(t)$ in $x_q(t)$ ob vzorčnih trenutkih zavzameta samo vrednosti iz njune abecede, zato ima tudi ekvivalentni signal $y_{eq}(t)$ ob vzorčnih trenutkih omejen nabor možnih vrednosti, ki jih lahko predstavimo v kompleksni ravnini, v tako imenovane



Slika 8.5 – Konstalacijski diagram 16QAM modulacije. Vsaka točka v konstalacijskem diagramu predstavlja par simbolov $(x_i[n], x_q[n])$. Realna vrednost je enaka simbolu v fazi, imaginarna pa simbolu v kvadraturi. Oddaljenost od izhodišča je amplituda, kot med absciso in kazalcem do točke pa faza, ki pripada posameznemu simbolu.

konstalacijskem diagramu. Na sliki 8.5 je prikazan primer konstalacijskega diagrama za 16QAM modulacijo.

Vsaka točka v diagramu predstavlja dvojico $(x_i[n], x_q[n])$ oddanih simbolov, oziroma 4 bite (binarne simbole), saj 16 različnih vrednosti lahko zapišemo s 4 biti. Realna vrednost je enaka $x_i[n]$ imaginarna pa $x_q[n]$. Oddaljenost od izhodišča je enaka amplitudi signala, ki pripada posameznemu simbolu, kot med kazalcem do točke in absciso pa je enaka fazi, ki pripada posameznemu simbolu. Ob prenosu se kazalec ekvivalentnega signala seli od točke do točke. Kadar so impulzi pravokotni, so ti prehodi skokoviti, kadar pa so impulzi oblikovani, pa so prehodi zvezni.

Kvadrat razdalje od izhodišča je proporcionalen energiji signala, ki jo potrebujemo za prenos enega simbola, srednja vrednost oddaljenosti pa moči signala, ki ga prenašamo. Ker bolj oddaljeni simboli doprinašajo več k moči signala, je smiselno preoblikovati konstalacijski diagram tako, da je čimbolj okrogle oblike. To je mogoče doseči z ustreznim kodiranjem. V praksi so v uporabi različne oblike konstalacijskih diagramov, predvsem pri večnivojski QAM modulaciji, kot na primer pri 256QAM.

8.3 Kotne modulacije

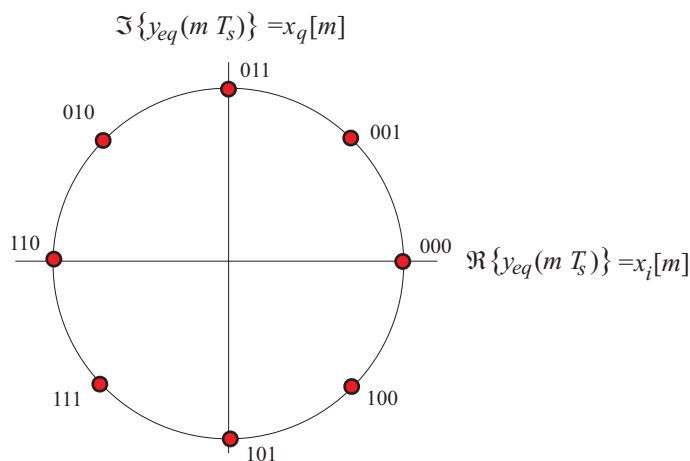
Pri digitalnih modulacijskih postopkih je še nekoliko težje ločevati med frekvenčno in fazno modulacijo kot pri analognih. Tako pri frekvenčni kot pri fazni modulaciji se namreč hkrati spreminjata trenutna frekvenca in trenutna faza signala. Al je podatek skrit v frekvenci ali fazi

signala pa ni odvisno samo od modulatorskega postopka temveč tudi od kanalskega kodiranja. Z ustreznim kodiranjem pred modulacijo lahko fazno modulacijo spremenimo v frekvenčno in obratno.

8.3.1 PSK

PSK (Phase shift keying) je modulacija s skokovitim preklapljanjem faze. Vsakemu simbolu iz abecede pripišemo določeno fazo med 0 in 2π . Fazno modulacijo lahko izvedemo kvadraturnim modulatorjem, kot je to prikazano na sliki 4.14. Na vejo v fazi vodimo kosinus faze na vejo v kvadraturi pa sinus faze.

Fazno modulacijo lahko prikažemo v konstalacijskem diagramu. Vse točke pri fazni modulatorji ležijo na krožnici in so med seboj enako oddaljene. Konstalacijski diagram za 8PSK je prikazan na sliki 8.6. Pri 8PSK so vhodni simboli osem nivojski. Z vsakim simbolom



Slika 8.6 – Konstalacijski diagram 8PSK. Vsaka točka v diagramu predstavlja en oddan simbol oziroma tri oddane bite.

torej prenašamo tri bite. Običajno pri uporabimo preslikavi bitov v simbole Grayevo kodo, kar pomeni, da se sosednja simbola razlikujeta samo v enem bitu. To pomeni, da napaka za eno fazo pomeni hkrati tudi napako enega bite, kar zmanjšuje pogostost bitne napake (BER).

Demodulacijo lahko izvedemo s QAM demodulatorjem ali faznim detektorjem. PSK izvedeno s kvadraturnim modulatorjem nekateri imenujejo tudi QPSK.

8.3.2 CPM

CPM (Continuous Phase Modulation) je modulatorski postopek enak kot PSK, le da se faza med vzorčnimi trenutki spreminja zvezno. Namesto pravokotnih impulzov v tem primeru v QAM modulatorju uporabimo impulze druge oblike. Če želimo, da se amplituda moduliranega signala ne spreminja oziroma, da je stalno enaka ena, mora biti izpolnjen pogoj:

$$x_i^2(t) + x_q^2(t) = 1 \quad (8.10)$$

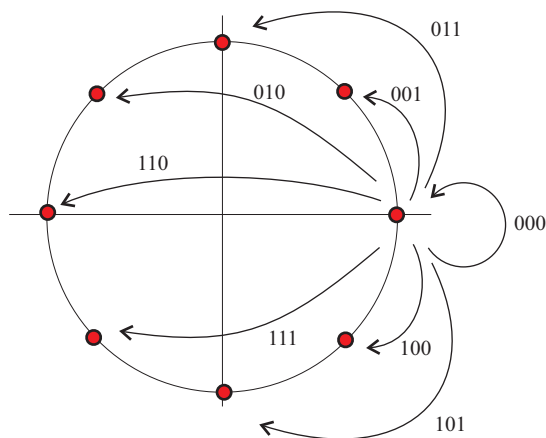
ki je vedno izpolnjen, če najprej ustvarimo fazni signal $\phi(t)$ in nato signala $x_i(t)$ in $x_q(t)$ po enačbah:

$$\begin{aligned}x_i(t) &= \cos \phi(t) \\x_q(t) &= \sin \phi(t)\end{aligned}\quad (8.11)$$

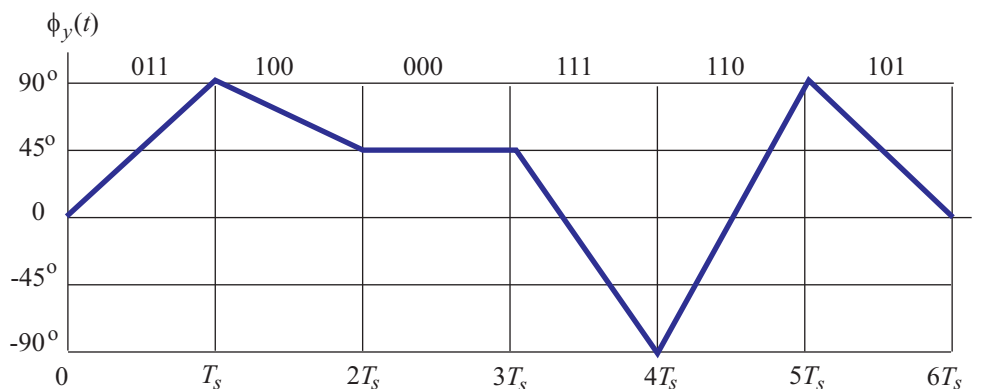
Konstalacijski diagram 8CPM je enak konstalacijskemu diagramu 8PSK. Razlika je le v prehodih med posameznimi fazami, ki v konstalacijskem diagramu niso vidne.

8.3.3 DPSK

DPSK (Differential Phase Shift Keying) je modulacija, kjer podatka ne prenašamo z absolutno fazo ampak je podatek v spremembi faze, kot je to prikazano na sliki 8.7.



Slika 8.7 – Pri DPSK prenašamo simbole s spremembo faze. Prednost je v tem, da pri sprejemu ni potrebno poznati absolutne faze signala. Zaznati je potrebno le spremembe v fazi.



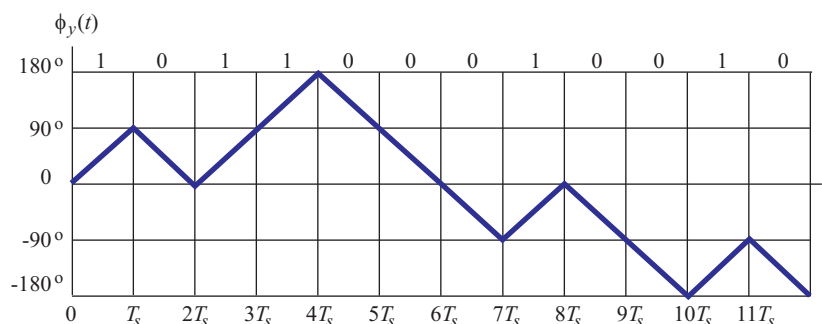
Slika 8.8 – Kadar so prehodi faze med vzorčnimi trenutki linearni, dobimo pri diferencialnem kodiranju za vsak vhodni simbol drugačen naklon, oziroma drugo frekvenco. Diferencialna fazna modulacija je zato v resnici frekvenčna modulacija oziroma FSK (Frequency Shift Keying).

Tudi pri DPSK je smiselno uporabiti Grayevo kodo, da zmanjšamo verjetnost bitne napake. Če so prehodi med posameznimi fazam niso skokoviti, temveč zvezno linearni, kot je to prikazano

na sliki 8.8, potem je DPSK v resnici frekvenčna modulacija. Odvod faze po času (naklon poteka faze) namreč predstavlja trenutno frekvenco. Vsak prehod faze ustreza določenemu simbolu, in torej tudi vsak simbol predstavlja svojo frekvenco. DPSK z linearnimi prehodi faze med simboli je torej v resnici FSK (Frequency Shift Keying) z zvezno fazo. Ker lahko frekvenco detektiramo brez da bi poznali nosilec, pri DPSK ni potrebna sinhronizacija nosilca.

8.3.4 MSK

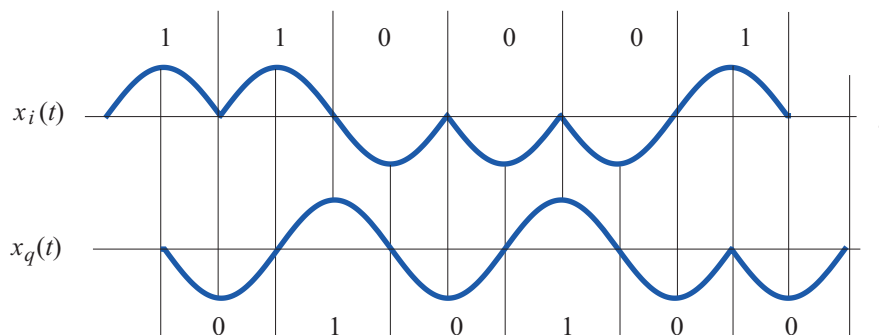
MSK (Minimum Shift Keying) je štirinivojska diferencialna fazna modulacija, kjer pa je preskok faze omejen na 90° . Logična 1 je kodirana s faznim skokom 90° in logična 0 s faznim skokom -90° , kot je to prikazano na sliki 8.9.



Slika 8.9 – Fazni potek MSK. Faza se spreminja linearno, tako da v signalu nastopata samo dve frekvenci. MSK je v bistvu 2FSK z modulacijskim indeksom $m = 0,5$.

Ker v faznem diagramu nastopata samo dva naklona, je to dvonivojska frekvenčna modulacija 2FSK. Pogosto imenujejo MSK tudi FFSK (Fast Frequency Shift Keying). Modulacijski indeks, ki je pri digitalni frekvenčni modulaciji definiran kot $m = \Delta f / f_s$, je pri MSK enak 0,5.

MSK lahko izvedemo s QAM modulatorjem. Kot smo že omenili, pri fazni modulaciji pripeljemo na vejo v fazi I kosinus faze in na vejo v kvadraturi Q sinus faze, kot je to prikazano na sliki 8.10.



Slika 8.10 – Signala, ki ju pripeljemo na QAM modulator pri MSK modulaciji. Vidimo, da so to za 90° zamaknjeni pol-sinusni impulzi.

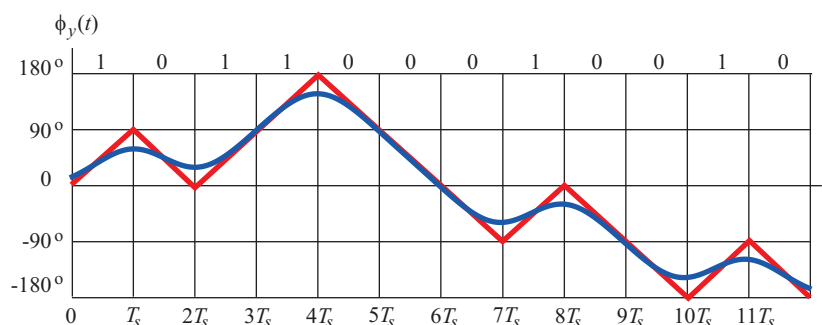
To omogoča tudi preprost izračun močnostnega spektra MSK. Za pol-sinusne impulze in s tem tudi za MSK je spekter prikazan že na desni strani slike 8.2.

8.3.5 GMSK

Lomljen potek faze na MSK signala, kot je prikazan na sliki 8.9 še vedno povzroča sevanje izven prenosnega pasu. Pri GMSK (Gaussian MSK) je fazni potek zglajen z Gaussovimi sitom, to je sitom z Gaussovo sistemsko funkcijo:

$$h(t) = e^{-(at/T_s)^2} \quad (8.12)$$

kjer konstanta a določa dolžino odziva oziroma pasovno širino sita. V GSM standardu je $a = 1,6$, kar ugodno vpliva na spekter signala, vendar povzroča tudi nekaj intersimbolne interference. Intersimbolno interferenco lahko vidimo na sliki 8.11, kjer je podana primerjava časovnega poteka faze GMSK signala (modra) s časovnim potekom faze MSK signala. MSK ima ob vzorčnih trenutkih pravilne faze, pri GMSK pa faze zaradi intersimbolne interference ob vzorčnih trenutkih odstopajo od teh vrednosti. Kljub temu pa sta obe frekvenci frekvenčne modulacije dobro ločljivi, kar je razvidno iz naklonov fazne karakteristike med posameznimi simboli.



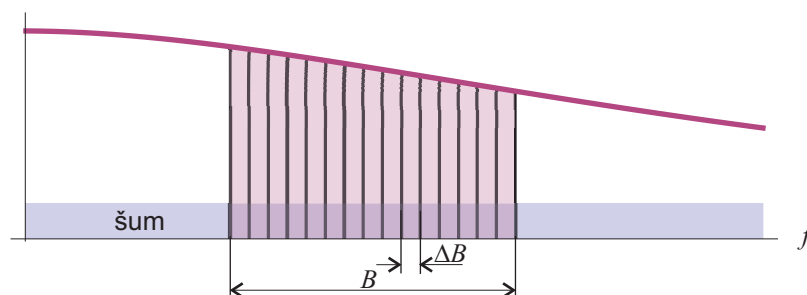
Slika 8.11 – Časovni potek faze GMSK signala (modra) in MSK signala (rdeča). Ob vzorčnih trenutkih zaradi intersimbolne interference GMSK faza odstopa od pravih vrednosti.

Modulacijo GMSK signala lahko izvedemo s QAM demodulatorjem. Ker je to frekvenčna modulacija, lahko demoduliramo tudi s frekvenčni demodulatorjem, pri katerem ne potrebujemo sinhronizacije nosilca.

8.4 Prenos z več nosilci

Kadar karakteristika kanala ni ravna, oziroma prekrita prevajalna funkcija ni konstantna, moramo, da ne bi prišlo do intersimbolne interference, karakteristiko kanala izravnati z izravnalnikom. Izravnava sicer zagotovi, da ni intersimbolne interference, vendar obenem pokvari razmerje med signalom in šumom (SNR).

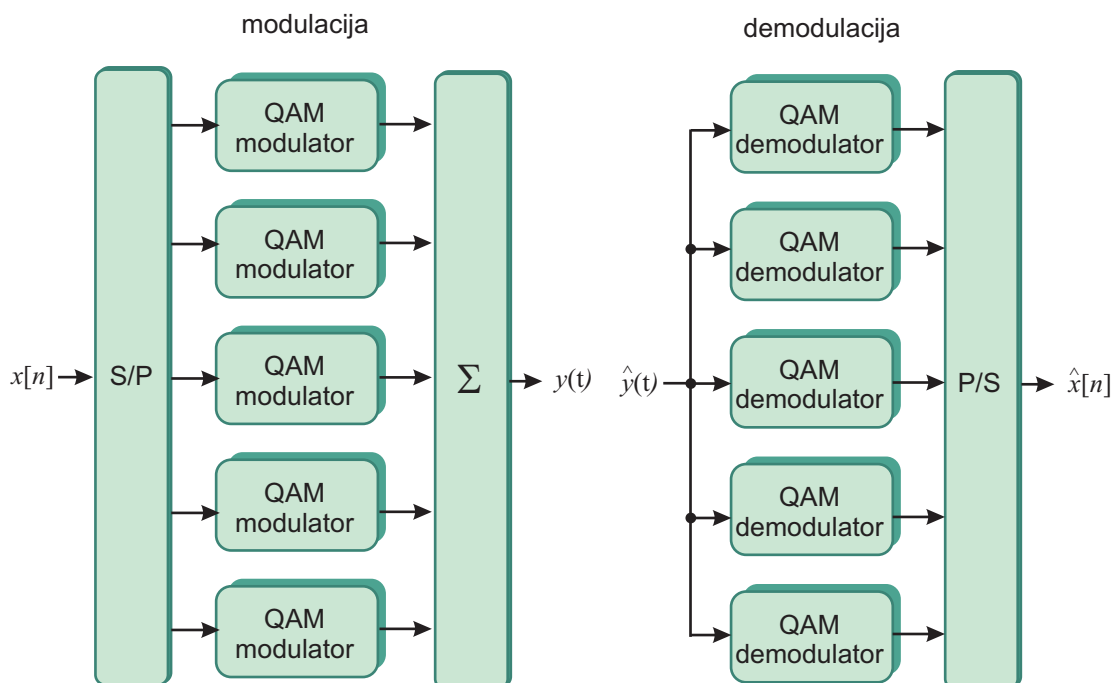
Da bi se izognili potrebi po izravnavi, lahko frekvenčni pas s pasovno širino B , ki smo ga namenili prenosu, razdelimo na več podkanalov s pasovno širino ΔB , kot je to prikazano na sliki 8.12. Ker je pasovna širina posameznega podkanala ΔB relativno majhna, se amplitudna karakteristika v enem podkanalu ne spreminja bistveno in je skoraj konstantna, zato ni potrebna izravnava.



Slika 8.12 – Pri prenosu z več nosilci razdelimo frekvenčno področje prenosa na veliko število podkanalov. Amplitudna karakteristika znotraj posameznega podkanala je skoraj konstantna. SNR je lahko v vsakem podkanalu drugačno.

Vhodni niz simbolov razdelimo na večje število paralelnih nizov in z vsakim nizom moduli-ramo nosilec, ki pripada ustreznemu podkanalu. Ker je v splošnem v vsakem podkanalu SNR drugačno, lahko, da bi boljše izkoristili zmogljivost posameznega podkanala, v vsakem podkanalu uporabimo različen modulacijski postopek z različnim številom nivojev. Zaradi velike spektralne učinkovitosti se v ta namen najpogosteje uporablja QAM z različnimi konstalacijskimi diagrami.

Načelna shema prenosa z več nosilci je prikazana na sliki 8.13. Na serijsko paralelnem pretvorniku vhodni niz simbolov razdelimo na več nizov z nižjo simbolno hitrostjo in vsakega vodimo na svoj modulator. Na sprejemni strani sprejeti signal razdelimo v več vej in v vsaki veji demoduliramo s svojim demodulatorjem. Za vsak kanal torej potrebujemo modulator na oddajni



Slika 8.13 – Načelna shema prenosa z več nosilci. Za vsak kanal potrebujemo svoj modulator na oddajni strani in demodulator na sprejemni strani. Tak način izvedbe za veliko število podkanalov ni primeren zaradi prevelike zahtevnosti in previsoke cene.

in demodulator na sprejemni strani. Poleg tega potrebujemo tako na oddajni kot na sprejemni

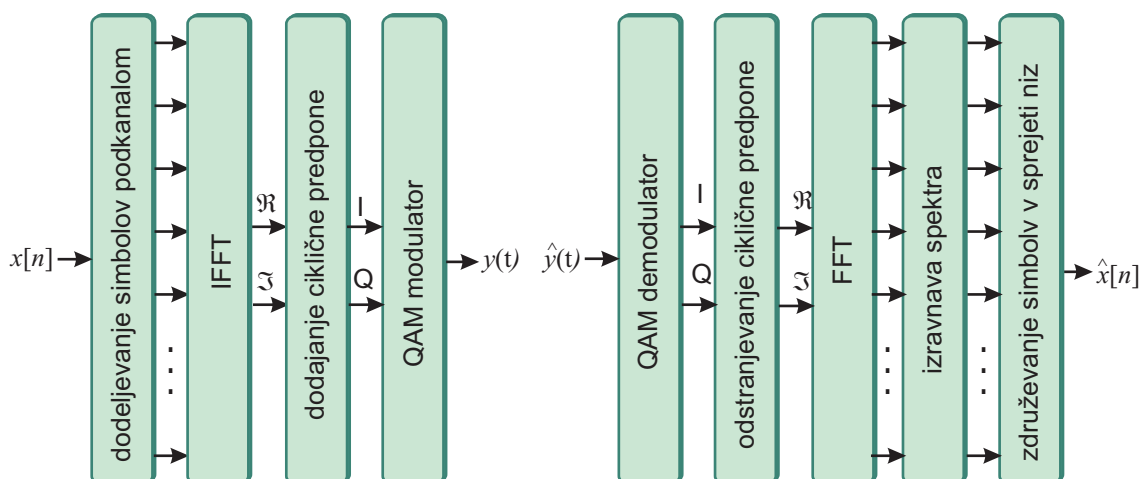
strani tudi izredno zahtevna ozkopasovna sita, ki med seboj ločujejo posamezne podkanale. Izvedba z analognimi vezji zato v praksi ni smiselna.

8.4.1 Digitalna izvedba prenosa z več nosilci

Pri digitalni izvedbi prenosa z več nosilci izvedemo postopka modulacije in demodulacije v frekvenčnem prostoru. Modulacijo izvedemo tako, da vsakemu nosilcu določimo ustrezno amplitudo in fazo, oziroma ustrezen realni in imaginarni del. Postopek ustreza QAM modulaciji nosilcev. Dobljeni spekter nato preslikamo v časovni prostor z uporabo inverznega diskretnega Fourierovega transformata (IDFT).

Pri demodulaciji sprejeti signal najprej preslikamo v frekvenčni prostor z uporabo diskretnega Fourierovega transformata (DFT) in nato odčitamo amplitude in faze oziroma realne in imaginarne dele sprejetih spektralnih komponent.

Postopek je računsko zelo učinkovit, ker lahko za izračun IDFT in DFT uporabimo hitre algoritme IFFT in FFT. Bločni diagram digitalne izvedbe modulacije in demodulacije z več nosilci je prikazan na sliki 8.14.



Slika 8.14 – Pri digitalni izvedbi prenosa z več nosilci ustvarimo spekter moduliranega signala v frekvenčnem prostoru in ga nato preslikamo v časovni prostor. Pri demodulaciji izvedemo obraten postopek. Dobljeni signal najprej preslikamo v frekvenčni prostor in iz dobljenega spektra odčitamo amplitude in faze nosilcev.

Da bi lahko uporabili FFT in IFFT mora biti število kanalov N cela potenca števila 2 (tipično 256, 512 ali 1024). Nosilce enakomerno porazdelimo v frekvenčno področje $0 < F < 1$ s frekvenčnim razmikom $\Delta F = 1/N$.

Vsakemu kanalu dodelimo določeno število binarnih simbolov (bitov) vhodnega binarnega niza. Koliko bitov dodelimo posameznemu kanalu je odvisno od SNR ki ga pričakujemo na tem kanalu. Bite, ki jih dodelimo posameznemu kanalu preslikamo v par večnivojskih simbolov oziroma v realni in imaginarni del kompleksnega simbola. Koliko nivojev ima ta, oziroma kakšna je njegova abeceda, je odvisno od števila bitov, ki smo jih dodelili posameznemu kanalu. Pri 16QAM modulaciji, na primer, predstavlja vsak simbol 4 bite, dva za realni in dva za imaginarni del. Kompleksen simbol, ki smo ga dobili na ta način, predstavlja kompleksno amplitudo

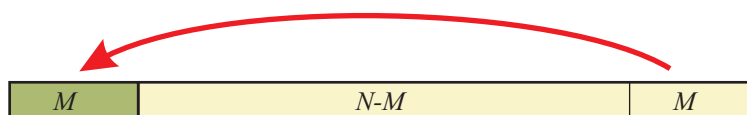
ustreznega nosilca.

Ker smo nosilce razvrstili enakomerno s frekvenčnim razmikom $1/N$, predstavljajo njihove kompleksne amplitude diskretni spekter periodičnega signala s periodo N . Z IFFT nato iz spektra signala izračunamo signal. Ker pri nastavljanju amplitud nismo upoštevali simetrije, ki velja za spekter realnih signalov, je signal, ki ga dobimo z IFFT v splošnem kompleksen. Dobljeni signal je enak vzorcem ekvivalentnega signala v osnovnem pasu, kot smo ga definirali v izrazu (8.9). Z vsakim IFFT pridobimo en blok N vzorcev ekvivalentnega signala. Z vsakim blokom prenesemo po en kompleksen simbol v vsakem kanalu. Trajanje simbola je torej enako N vzorcem ekvivalentnega signala.

Ker smo blok N vzorcev izračunali z IFFT, smo implicitno predpostavili, da se izven tega bloka signal periodično ponavlja, kar pa v resnici ne drži. Z naslednjim blokom bomo namreč prenašali druge simbole in bo zato tudi signal drugačen. Signal, ki ga prenašamo ni periodičen.

V zvezi s tem se postavi vprašanje, ali in v kolikšni meri lahko pri sprejemu določenega bloka vplivajo sosednji bloki, to je bloki, ki so bili oddani prej in bloki, ki so bili oddani kasneje. Ker je dolžina enega bloka enaka dolžini enega oddanega simbola, predstavlja vpliv med bloki intersimbolno interferenco. Načeloma lahko na trenutni blok vplivajo le bloki, ki so jih oddali prej (kavzalna intersimbolna interferenca). Kako dolgo lahko predhodni bloki motijo sprejem trenutnega bloka je odvisno od trajanja odziva kanala. Ko odziv kanala izzveni, prej oddani bloki nimajo več vpliva na sprejem trenutnega bloka.

Če bi bil signal, ki ga oddajamo periodičen, kot smo predpostavili, ko smo naredili IFFT, bi bil sprejeti blok pravilen. Ker na sprejem vpliva le signal, ki je manj kot za dolžino odziva kanala oddaljen od trenutnega bloka, zadošča, da je signal periodičen zgolj v tem delu. Zato vsakemu bloku dodamo ciklično predpono. M vzorcev s konca bloka kopiramo pred začetek bloka, kot je to prikazano na sliki 8.15.



Slika 8.15 – Da bi se pri sprejemu sprejeti signal obnašal, kot da bi bil periodičen, kopiramo M vzorcev bloka pred začetek bloka. Število vzorcev M mora ustrezati dolžini odziva kanala, da ne pride do intersimbolne interference.

To nekoliko zmanjša prenosno hitrost, saj moramo v vsakem bloku prenašati M odvečnih simbolov. Vendar to zmanjšanje ni bistveno kadar je N velik in mnogo večji od M .

Tako dobljeni signal nato vodimo na QAM modulator, realni del na I vejo in imaginarni del na Q vejo modulatorja. Na sprejemni strani sprejeti signal najprej demoduliramo s QAM demodulatorjem, nato odstranimo ciklično predpono in izračunamo FFT. Vsakemu nosilcu se pri prenosu lahko spremenita amplituda in faza. To pomeni, da so spektralne komponente, ki predstavljajo posamezne nosilce množene vsaka s svojo kompleksno konstanto. Pri sprejemu spekter zopet izravnamo, tako, da kompleksne amplitude spektra množimo z inverznimi vrednostmi teh konstant. Taka izravnava ne poslabša SNR v podkanalih. Dobljene simbole nato zopet združimo v sprejeti niz z obratnim postopkom, kot smo ga uporabili pri dodeljevanju bitov

posameznim kanalom.

Prenos z več nosilci je standardiziran predvsem v dveh različicah, to je v OFDM (Orthogonal Frequency Division Multiplex) ki se uporablja v sistemu WiMax in DMT (Discrete Multi-Tone), ki se uporablja v digitalnem radiu. V teh standardih je predpisanih več različnih modulacijskih postopkov oziroma konstalacijskih diagramov za posamezne podkanale glede na razmerje med kanalom in šumom v posameznih podkanalih. Določen je tudi način merjenja karakteristike kanala in način dogovarjanja med oddajnikom in sprejemnikom o shemi dodeljevanja bitov posameznim podkanalom.

8.5 Prenos z razširjenim spektrom

Pri prenosu z razširjenim spektrom (SS - Spread Spectrum) namenimo prenose bistveno večjo pasovno širino, kot je minimalno potrebna za prenos z določeno simbolno hitrostjo. Čeprav potrebujemo za prenos z razširjenim spektrom bistveno večjo pasovno širino, kot bi bilo potrebno, pa ima tak prenos določene prednosti, o katerih bo več govora v nadaljevanju. Speker lahko razširimo predvsem na dva načina, to je s frekvenčnim skakanjem in s časovnim skakanjem.

8.5.1 Frekvenčno skakanje

Pri frekvenčnem skakanju (FH/SS – Frequency Hopping Spread Spectrum) spreminjamo frekvenco nosilca po nekem določenem zaporedju, ki ga imenujemo tudi koda. Kodo morata poznati tako modulator kot demodulator. Načeloma lahko pri tem uporabimo poljuben modulacijski postopek, vendar postopki, ki potrebujejo sinhronizacijo nosilca pri demodulaciji, niso najprimernejši. Zato se v povezavi s frekvenčnim skakanjem najpogosteje uporabljata FSK in PAM.

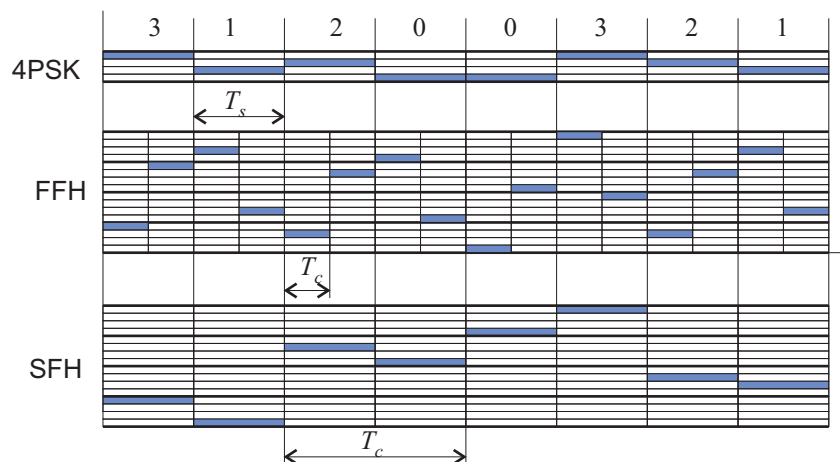
Ločimo počasno frekvenčno skakanje (SFH - Slow Frequency Hopping) in hitro frekvenčno skakanje (FFH - Fast Frequency Hopping).

Pri počasnem frekvenčnem skakanju je čas med dvema skokoma T_c (ang. chip time) večji od časa med dvema simboloma T_s , kar pomeni, da na eni frekvenci prenesemo več simbolov, preden skočimo na drugo frekvenco.

Pri hitrem skakanju je T_c manjši od T_s , kar pomeni, da med prenosom enega samega simbola naredimo več frekvenčnih skokov. Izvedba hitrega frekvenčnega skakanja je mnogo zahtevnejša od izvedbe počasnega skakanja.

Primer, kako se spreminja frekvenca pri uporabi 4FSK modulacije pri hitrem in počasnem skakanju, je shematsko prikazan na sliki 8.16.

Omeniti velja tudi zelo počasno frekvenčno skakanje, kjer je T_c veliko večji od T_s . To pomeni, da prenesemo veliko količino podatkov na eni frekvenci, preden preklopimo na drugo frekvenco. Pri zelo počasnem skakanju ne gre za pravo razširjanje spektra, saj je prenos med dvema skokoma ozkopasoven. Zelo počasno skakanje se uporablja v GSM mobilnem omrežju za izogibanje frekvenčnim luknjam, ki so posledica frekvenčnega presiha (ang. frequency fading) zaradi prenosa po več poteh, in ozkopasovnim motnjam v mobilnem prenosnem kanalu.



Slika 8.16 – Časovne spremembe frekvence pri 4PSK, 4PSK/SFH in 4PSK/FFH. Pri počasnem frekvenčnem skakanju se prenese več simbolov preden se spremeni frekvenca nosilca, vendar se znotraj tega frekvence spreminja zaradi frekvenčne modulacije. Pri hitrem skakanju frekvenca spreminja med prenosom enega samega simbola.

8.5.2 Neposredno razširjanje s kodnim zaporedjem

Neposredno razširjanje s kodnim zaporedjem (DS/SS Direct Sequence Spread Spectrum) je pravzaprav amplitudna modulacija psevdonaključnega signala. Namesto da bi z modulatorskim signalom $x(t)$ modulirali amplitudo harmoničnega nosilca $\cos(2\pi f_{vm}t)$ tu z njim množimo psevdonaključni signal $c(t)$, ki ga imenujemo tudi koda.

$$y(t) = x(t)c(t) \quad (8.13)$$

Koda $c(t)$ je lahko poljubna širokopasovna naključna signal, ki ga morata poznati tako modulator kot demodulator. Običajno se za ta namen uporablja niz impulzov, ki ga generiramo iz psevdonaključnega niza števil: $c[m]$:

$$c(t) = \sum_m c[m]g_c(t - mT_c) \quad (8.14)$$

kjer predstavlja $g_c(t)$ obliko impulza, T_c pa je časovni razmik med impulzi (ang. chip time), ki je bistveno krajši od simbolnega časa T_s . Zaradi preprostejšie izvedbe je niz $c(t)$ običajno bipolaren ($c[n] \in \{-1, 1\}$), impulzi $g_c(t)$ pa pravokotni impulzi dolžine T_c . Iz enakega razloga se tudi podatkovni simboli prenašajo z bipolarnim nizom pravokotnih impulzov $g_s(t)$ dolžine T_s :

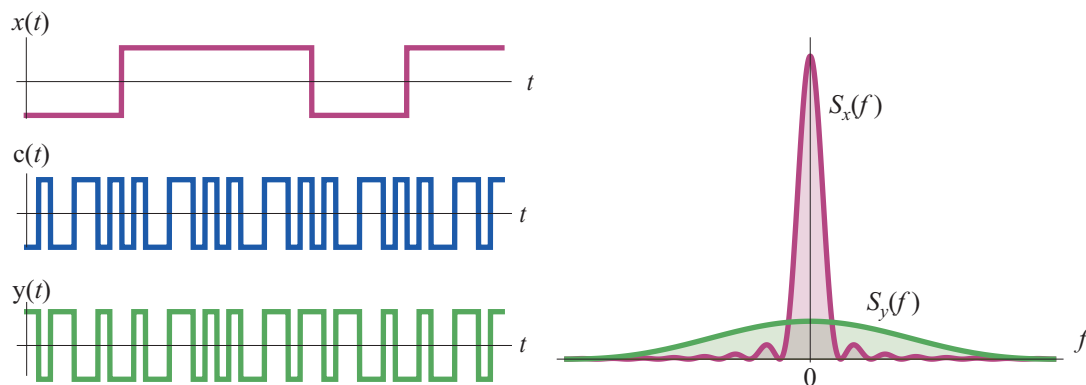
$$x(t) = \sum_n x[n]g_s(t - nT_s) \quad (8.15)$$

Razmerje med dolžino simbola T_s in dolžino impulza kode T_c imenujemo tudi kodno ojačenje G_c :

$$G_c = \frac{T_s}{T_c} \quad (8.16)$$

S tem faktorjem se tudi razširi močnostni spekter.

Primeri signalov $x(t)$, $c(t)$ in $y(t)$ ter močnostna spektra ozkopasovnega in razširjenega signala pri $G_c = 8$ so prikazani na sliki 8.17. Močnostni spekter kode $S_c(f)$ je enak močnostnemu spektru razširjenega signala $S_y(f)$ zato ni posebej prikazan. V praksi se uporablja bistveno večja kodna ojačenja. Tako majhno kodno ojačenje je na sliki samo zato, da je slika bolj nazorna.



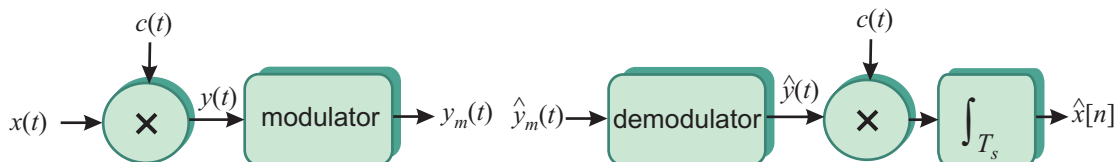
Slika 8.17 – Levo: modulacijski signal $x(t)$, koda $c(t)$ in razširjen signal $y(t)$ pri kodnem ojačenju $G_c = 8$. Desno: Močnostni spekter modulacijskega signala $S_x(f)$ in razširjenega signala $S_y(f)$. Močnostni spekter kode $S_c(f)$ je enak približno močnostnemu spektru razširjenega signala, zato ni prikazan.

Pri neposrednem razširjanju s kodnim zaporedjem je razširjeni signal še vedno v osnovnem pasu. Zato, da ga prestavimo v višjo frekvenčno lego, lahko uporabimo poljuben modulacijski postopek.

Razširjeni signal demoduliramo oziroma zožimo s korelacijskim sprejemnikom. Korelacijski sprejemnik naredi korelacijo med sprejetim signalom $\hat{y}(t)$ in kodo $c(t)$ preko enega simbolnega intervala, to je za vsak simbol posebej:

$$\hat{x}[n] = \int_{nT_s}^{(n+1)T_s} c(t)\hat{y}(t)dt \quad (8.17)$$

Sprejeti signal torej pri sprejemu primerjamo z znano kodo. Če je znana koda prisotna v signalu, potem je korelacija velika, bodisi pozitivna (simbol +1) ali negativna (simbol -1). Shematsko je celoten sistem prikazan na sliki 8.18.



Slika 8.18 – Razširjanje spektra in sprejem razširjenega signala. Pri oddaji podatkovni signal $x(t)$ množimo s kodo $c(t)$ in nato še moduliramo, da prestavimo v ustrezno frekvenčno področje. Pri sprejemu sprejeti signal najprej demoduliramo, nato pa naredimo še korelacijo z znano kodo $c(t)$.

Da bi pri sprejemu lahko naredili korelacijo med kodo in sprejetim signalom čez en simbol, morata biti oddajnik in sprejemnik sinhronizirana. Potrebi po sinhronizaciji se lahko izognemo, če pri sprejemu namesto korelacije delamo korelacijsko funkcijo dolžine enega simbolnega inter-

vala. Ko sta koda v sprejetem signalu in koda v sprejemniku časovno poravnani, je korelacijska funkcija maksimalna.

8.5.3 Prednosti razširjanja spektra

Čeprav potrebujemo za prenos z razširjenim spektrom bistveno večji frekvenčni pas kot pri ozkopasovnem prenosu, pa je prenos z razširjenim spektrom smislen zaradi prednosti, ki jih ima pred ozkopasovnim prenosom. Tu omenimo zgolj prednosti, ki jih ima tak prenos v vojaških komunikacijah, kjer se je najprej uveljavil, v naslednjem poglavju, kjer bomo govorili o sodostopu do skupnega prenosnega medija, pa bomo opisali tudi prednosti razširjanja spektra v civilnih komunikacijah, predvsem v celičnih omrežjih mobilne telefonije.

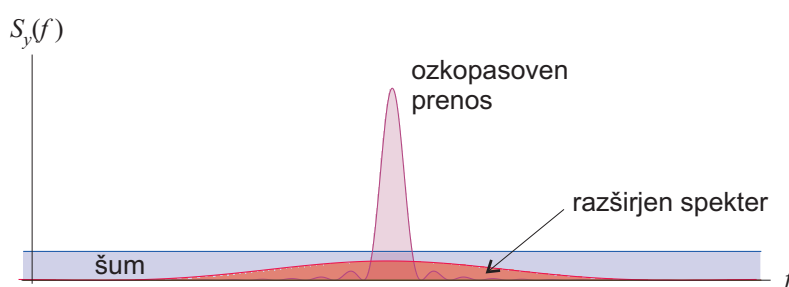
V vojaških komunikacijah so prednosti razširjanja spektra predvsem zasebnost, skritost, odpornost na ozkopasovne motnje in odpornost na frekvenčni presih zaradi prenosa po več poteh.

8.5.3.1 Zasebnost

Da bi lahko sprejemali signal, ki se prenaša z razširjenim spektrom, moramo poznati kodo, s katero je bil spekter razširjen. Ta koda je lahko psevdonaključno zaporedje, ki ga nasprotnik ne pozna, zato ne more prisluškovati komunikaciji. V vojaških komunikacijah se sicer poleg te zaščite zasebnost dodatno ščiti še s šifriranjem sporočil.

8.5.3.2 Skritost

Moč signala je enaka ploščini pod funkcijo gostote močnostnega spektra $S_y(f)$. Pri prenosu z enako močjo je zato pri razširjenem spektru gostota močnostnega spektra bistveno nižja kot pri ozkopasovnem prenosu in lahko pade tudi pod nivo šuma, kot je to shematsko prikazano na sliki 8.19. Signal je skrit v šumu, tako da nasprotnik sploh ne more zaznati, da komunikacija poteka.

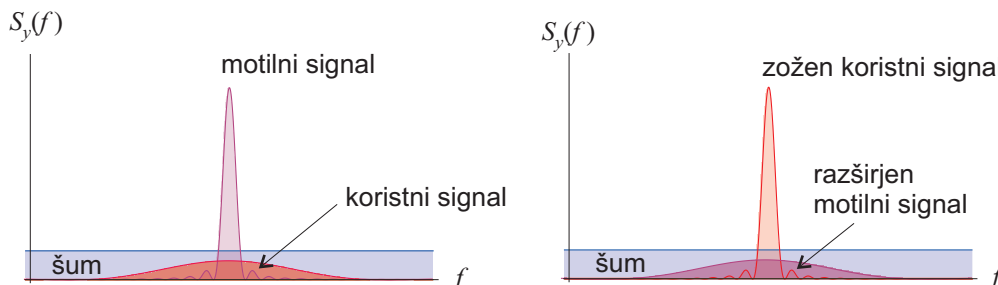


Slika 8.19 – Pri enaki moči signala je gostota močnostnega spektra pri prenosu z razširjenim spektrom bistveno nižja kot pri ozkopasovnem prenosu in je lahko pod nivojem šuma.

8.5.3.3 Odpornost proti ozkopasovnim motnjam.

Nasprotnik lahko ozkopasovne komunikacije moti z močnim signalom na frekvenci, kjer poteka komunikacija (ang. narrow band jamming). Pri prenosu z razširjenim spektrom pri sprejemu

spekter signala zopet zožimo, pri tem pa se razširi ozkopasovna motnja in zato ne onemogoča sprejema, kot je to prikazano na sliki 8.20. Če bi želel nasprotnik motiti tako komunikacijo,



Slika 8.20 – Odpornost proti ozkopasovnim motnjam. Pri sprejemu se signal zoži, ozkopasovna motnja pa se razširi in ne onemogoča sprejema

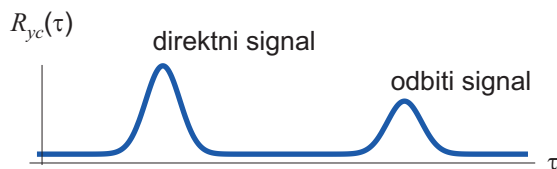
bi moral oddajati z veliko močjo na celotnem področju razširjenega signala, kar pa pri velikem razširjanju ni izvedljivo, poleg tega pa bi motilo tudi njegove lastne komunikacije.

8.5.3.4 Odpornost na frekvenčni presih

V radijskih komunikacijah poteka prenos pogosto po več poteh. Na sprejemniku poleg direktnega signala sprejemamo tudi signale, ki so se odbili od različnih ovir, kot na primer hribov ali večjih stavb. Odbiti signali so glede na direktni signal zakasnjeni. Pri ozkopasovni modulaciji sta lahko direktni in odbiti signal v protifazi in se odštejeta. Govorimo o tako imenovanem frekvenčnem presihi (ang. frequency fading), ki lahko onemogoči sprejem signala.

Pri prenosu z razširjenim spektrom zavzema frekvenčna luknja, ki nastane zaradi frekvenčnega presiha, le majhen del celotne pasovne širine signala in zato ne onemogoča prenosa.

Pri sprejemu razširjenega signala naredimo korelacijsko funkcijo sprejetega signala in znane kode. Koda nastopa tako v direktnem kot tudi v odbitem signalu. Ker je odbiti signal zakasnen glede na direktni signal, nastopi v korelacijski funkciji še en maksimum, ki ustreza odbitemu signalu, kot je to prikazano na sliki 8.21. Direktni signal tako pri sprejemu preprosto ločimo



Slika 8.21 – Korelacijska funkcija pri sprejemu direktnega in odbitega signala. Prvi maksimum nastopi v času, ko pride na sprejemnik direktni signal, drugi pa, ko pride odbiti. Pri sprejemu je možno izkoristiti oba signala.

od odbitega signala, tako, da odbiti signal ne moti sprejema. Še več, pri sprejemu lahko konstruktivno izkoristimo oba signala in na ta način še izboljšamo sprejem. Sprejemnik, ki izkorišča direktni in enega ali več odbitih signalov imenujemo RAKE sprejemnik. Ime RAKE izvira iz angleške besede rake – grablje.

Poglavje obravnava različne načine sodostopa do skupne prenosne medija: frekvenčno deljen sodostop, časovno deljen sodostop, kodno deljen sodostop, sodostop na osnovi prenosa z več nosilci in naključni sodostop.

Pogosto si mora pri prenosu več uporabnikov deliti skupen prenosni medij. Najbolj značilno je za prenos po praznem prostoru oziroma atmosferi. Več uporabnikov pa si lahko deli tudi druge prenosne medije, kadar njihova zmogljivost presega potrebe posameznega uporabnika. To velja tako za žične kot optične povezave.

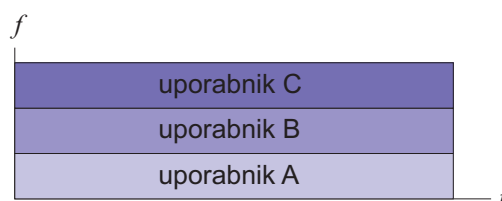
Pri sodostopu do skupnega medija moramo zagotoviti, da je na sprejemni strani možno ločiti signale, ki pripadajo različnim uporabnikom. Z drugimi besedami povedano to pomeni, da moramo ustvariti med seboj ločene prenosne kanale, ki pripadajo posameznim uporabnikom, na tak način, da v čim večji meri zmanjšamo njihov medsebojni vpliv. Število kanalov mora biti enako številu uporabnikov, ki oddajajo na skupnem prenosnem mediju. Uporabnikov, ki sprejemajo posamezen kanal pa je lahko tudi več. Kadar več uporabnikov sprejema isti kanal govorimo o razširjeni oddaji (ang. broadcasting) ki je značilna za radijske in televizijske programe.

O sodostopu govorimo tudi, kadar sta vključena dva uporabnika v dvosmerni (duplex) komunikaciji, saj je potrebno vsakemu dodeliti kanal, na katerem lahko on oddaja, drugi uporabnik pa sprejema.

Posamezne kanale lahko med seboj ločimo po frekvenci, po času ali po neki kodi, ki je značilna za posameznega uporabnika.

9.1 Frekvenčno deljen sodostop

Pri frekvenčno deljenem sodostopu (FDMA – Frequency Division Multiple Access) vsakemu uporabniku dodelimo določen frekvenčni pas, v katerem lahko oddaja, kot je to prikazano na sliki 9.1. Frekvenčni pas je posameznemu uporabniku dodeljen za ves čas komunikacije. V



Slika 9.1 – Pri frekvenčno deljenem sodostopu so uporabniki med seboj deljeni po frekvenčnem pasu, v katerem oddajajo. Pri tem lahko uporabljajo različne modulacijske postopke v kombinaciji z mešanjem.

sistemih, kjer je več uporabnikov kot je na voljo frekvenčnih kanalov, je zato potreben nek protokol za vzpostavljjanje zveze oziroma za zaseganje kanala na začetku in za rušenje zveze oziroma sproščanje kanala na koncu komunikacije.

Pri frekvenčnem deljenem sodostopu je pomembno, da spekter signala posameznega uporabnika ne presega pasovne širine kanala, ki mu je dodeljen, oziroma da je sevanje izven kanala čim manjše. Spekter, ki sega v sosednje kanale namreč moti prenos v sosednjih kanalih. Poleg tega želimo, da so kanali čim ožji in čim bližje skupaj, da na ta način kar najbolje izkoristimo prenosni medij. Za dobro izkoriščanje prenosnega medija so zato potrebni spektralno učinkoviti modulacijski postopki.

Kljub uporabi učinkovitih modulacijskih postopkov pa se ne moremo popolnoma izogniti sevanju izven kanala, saj bi zato potrebovali idealna oddajna sita. Zato morajo biti posamezni kanali med seboj nekoliko razmaknjeni, kar nekoliko zmanjša izkoriščenost prenosne poti. Pasovna širina B , ki je potrebna za K kanalov je torej enaka

$$B = K(B_k + B_v) = KB_k \left(1 + \frac{B_v}{B_k}\right) \quad (9.1)$$

Kjer je B_k pasovna širina enega kanala in B_v varnostni frekvenčni razmik me kanali, ki omogoča filtriranje in preprečuje sevanje izven kanala.

Ker lahko posamezne kanale pri sprejemu ločimo z ustreznimi pasovno prepustnimi siti, pri frekvenčno deljenem sodostopu ni potrebna sinhronizacija med uporabniki. Prav zato je predstavlja frekvenčno deljen sodostop tudi primarni način delitve med večje uporabnike, kot na primer, ponudnike telekomunikacijskih storitev, ki znotraj njim dodeljenega frekvenčnega pasu ustvarijo manjše prenosne kanale za individualne uporabnike.

Delitev med frekvenčnega spektra med velike uporabnike je še posebej pomembna pri brezžičnih komunikacijah, kjer si vsi uporabniki delimo skupno atmosfero. Frekvenčni spekter v atmosferi zato šteujemo za nacionalno bogastvo, koncesije za njegovo uporabo pa podeljuje država.

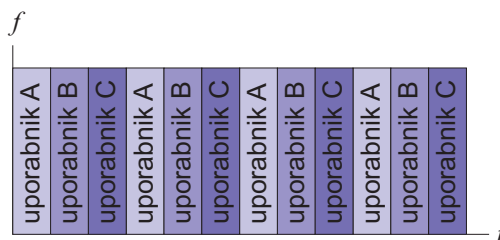
Uporaba spektra mora biti usklajena tudi na meddržavnem in globalnem nivoju, za kar skrbijo posebne organizacije.

9.2 Časovno deljen sodostop

Pri časovno deljenem sodostopu (TDMA – Time Division Multiple Access) imajo uporabniki na voljo celotno pasovno širino (dodeljeno s primarno delitvijo frekvenčnih pasov), vsakemu pa je dodeljena točno določena časovna rezina, v kateri lahko oddaja svoj blok podatkov, kot je to prikazano na sliki 9.2.

Da bi lahko oddajali v pravilni časovni rezini in da bi lahko sprejemali pravilno časovno rezino, morajo biti pri časovno deljenem sodostopu uporabniki med seboj sinhronizirani. Ker na prihaja na prenosni poti do zakasnitev, ki so odvisne od oddaljenosti uporabnikov, je potrebno med uporabnikom dodeljene časovne intervale dodati varnostne intervale. Čas T ki ga potrebuje K uporabnikov za oddajo bloka podatkov je potem enak:

$$T = K(T_k + T_v) = KT_k \left(1 + \frac{T_v}{T_k}\right) \quad (9.2)$$



Slika 9.2 – Pri časovno deljenem sodostopu ima vsak uporabnik na voljo celoten frekvenčni pas, vendar le v točno določeni časovni rezini. Uporabniki morajo biti zato med seboj sinhronizirani.

kjer je T_k čas, ki je na voljo za prenos enega bloka podatkov in T_v varnostni interval. Razmerje med časom, ki ga ima na voljo uporabnik v frekvenčno deljenem sodostopu, kjer ima za prenos bloka na voljo ves čas T in časom, ki ga ima na voljo v časovnem deljenem sodostopu T_k je torej enako:

$$\frac{T}{T_k} = K \left(1 + \frac{T_v}{T_k} \right) \quad (9.3)$$

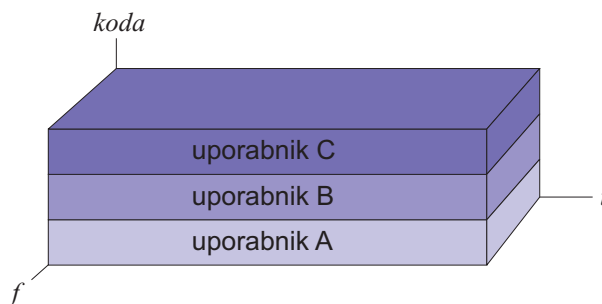
Če želimo oddajati z enako efektivno podatkovno hitrostjo pri frekvenčno deljenem sodostopu, je potreba za ta faktor povečati podatkovno hitrost znotraj časovne rezine, kar pomeni, da potrebujemo za prenos za tudi toliko večjo pasovno širino:

$$B = K B_k \left(1 + \frac{T_v}{T_k} \right) \quad (9.4)$$

kjer je B_k pasovna širina, ki bi jo potrebovali, če bi imeli na voljo ves čas za prenos. Vidimo, da je pasovna širina, ki jo potrebujemo za prenos K kanalov pri časovno deljenem sodostopu približno enaka kot pri frekvenčno deljenem sodostopu.

9.3 Kodno deljen sodostop

Pri kodno deljenem sodostopu (CDMA - Code Division Multiple Access) poteka prenos z razširjenim spektrom, tako da imajo ves čas prenosa vsi uporabniki na voljo celotno pasovno širino. Med seboj se ločujejo po kodi, ki je dodeljena vsakemu uporabniku, kot je to prikazano na sliki 9.3.



Slika 9.3 – Pri kodno deljenem sodostopu se uporabniki ločijo med seboj po kodah.

Kodno deljen sodostop ima pred časovno in frekvenčno deljenima sodostopoma kar nekaj prednosti, predvsem v celičnih omrežjih mobilnih komunikacij. To je tudi razlog, da se v tretji

generacija (3G) mobilnih komunikacij (UMTS, CDMA2000) uporablja kodni sodostop. S kodnim sodostopom je namreč mogoče bistveno povečati kapaciteto omrežja, to je število uporabnikov, ki lahko hkrati komunicirajo v dodeljenem frekvenčnem področju. Pleg tega ima razširjanje spektra še druge prednosti, ki smo jih že omenjali v prejšnjem poglavju.

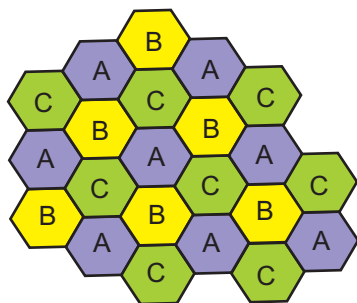
Prednosti kodnega sodostopa v celičnih omrežjih, ki omogočata povečanje kapacitete sistema, sta predvsem v možnosti ponovne uporaba frekvenc in dinamičnega zaseganja kanala. Poleg tega omogoča uporaba RAKE sprejemnika tudi mehko prehajanje med celicami.

9.3.1 Ponovna uporaba frekvenc

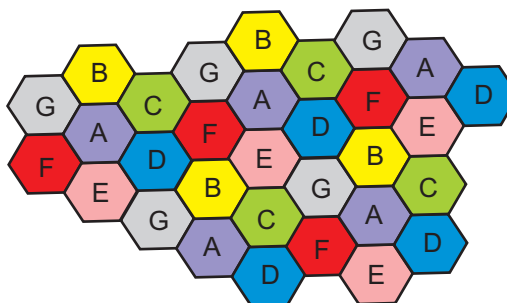
V celičnem omrežju je celotno omrežje razdeljeno na celice. Vsaka celica ima svojo bazno postajo z oddajno in sprejemno anteno. Uporabniki lahko prehajajo med celicami in se povežejo na bazno postajo, ki ima najmočnejši signal.

Pri frekvenčno deljenem sodostopu morajo, da se uporabniki ne bi motili med seboj, sosednje celice delovati v različne frekvenčnih pasovih. Da bi to lahko dosegli, je potrebno celotno frekvenčno področje razdeliti na tri do sedem frekvenčnih pasov – faktor ponovitve frekvence 3 do 7 – kot je to prikazano na sliki 9.4. Z večanjem faktorja ponovitve se večja oddaljenost med celicami z isto frekvenco in medcelična interferenca, vendar se hkrati manjša izkoriščenost frekvenčnega področja in s tem kapaciteta sistema.

faktor ponovitve frekvence = 3



faktor ponovitve frekvence = 7



Slika 9.4 – Frekvenčno planiranje v frekvenčno deljenem sodostopu. Z večanjem faktorja ponovitve (števila frekvenc) se povečuje razdalja med celicami z isto frekvenco in s tem medcelična interferenca.

V praksi niso vse celice enako velike, kar otežkoči frekvenčno planiranje. Včasih je potrebno, zaradi povečanega števila uporabnikov zmanjšati velikost celic, oziroma med obstoječe celice vriniti nove celice. Pri tem je potrebno na novo narediti celoten frekvenčni plan.

Pri kodno deljenem sodostopu se uporabniki med seboj ločijo po kodi. Če so kode med seboj nekorelirane se uporabniki med seboj ne motijo, čeprav vsi komunicirajo v istem, razširjenem frekvenčnem pasu. Zato lahko v vsaki celici izkoristimo celotno frekvenčno področje, kar pomeni, da je frekvenčni pas, ki je na voljo v posamezni celici tri do sedemkrat širši kot pri frekvenčno deljenem sodostopu.

Ker so, kot bo govora v nadaljevanju, kode med seboj zgolj približno nekorelirane, se uporabniki med seboj deloma motijo. Zaradi teh motenj lahko v določenem frekvenčnem pasu hkrati

komunicira le približno pol toliko uporabnikov kot pri frekvenčno deljenem sodostopu. Ker pa je frekvenčni pas 3 do 7 krat širši, je kapaciteta v kodnem sodostopu zaradi ponovne uporabe frekvenc lahko za faktor 1,5 do 3,5 krat večja od kapacitete pri frekvenčno ali časovno deljenem sodostopu.

9.3.2 Dinamično zaseganje kanala

Pri kodno deljenem sodostopu so uporabniki med seboj ločeni po kodah. Vsakemu uporabniku je dodeljena svoja koda. Pri komunikaciji zato zanj ni potrebno rezervirati posebnega frekvenčnega kanala, kot pri frekvenčno deljenem sodostopu oziroma časovne rezine, kot pri časovno deljenem sodostopu. Ker kode niso med seboj popolnoma nekorelirane, se uporabniki med seboj motijo, kar omejuje število uporabnikov, ki lahko hkrati komunicirajo.

Pri pogovoru po telefonu je posamezen uporabnik običajno približno polovico časa tiho, to je v času, ko govori njegov sogovornik. Pri frekvenčno ali časovno deljenem sodostopu tudi ta čas zaseda prenosni kanal, tako, da ga ne morejo izkoriščati drugi uporabniki, v kodnem sodostopu pa v času, ko ne govori, tudi ne moti drugih uporabnikov. Lahko rečemo, da je v času tišine sprosti prenosni kanal. To imenujemo dinamično zaseganje kanala, saj uporabnik zasega kanal le takrat, ko ga dejansko uporablja. Zaradi dinamičnega zaseganja kanala se pri prenosu govora poveča kapaciteta omrežja približno za faktor 2.

9.3.3 Mehko prehajanje med celicami

V prejšnjem poglavju smo omenili, da lahko z uporabo RAKE sprejemnika ločimo direktni in odbiti signal, ki pride do sprejemnika zakasnen in ju konstruktivno seštejemo pri sprejemu in se izognemo problemu frekvenčnega presiha zaradi prenosa po več poteh.

Na enak način lahko pri sprejemu ločim, kadar isti signal prihaja iz dveh baznih postaj z različni časovnim zamikom, kar omogoča mehko prehajanje med celicami. Ko se uporabnik približa meji med celicami, to je, ko ga zazna bazna postaja v sosednji celici, začne ta oddajati isti signal, kot bazna postaja v celici, v kateri se nahaja uporabnik. Z RAKE sprejemnikom lahko uporabnik hkrati sprejema signala obeh celic. Ko se približuje bazni postaji nove celice, se njen signal povečuje, signal prejšnje celice pa zmanjšuje, dokler, ko je od nje dovolj oddaljen, originalna celica ne preneha oddajati signala za tega uporabnika.

Ker uporabnik ne zamenja prenosnega kanala pri prehodu med celicami, ni potrebe, da bi se prenos med prehodom med celicami prekinil in ponovno vzpostavil na drugem kanalu. Zato tak prehod med celicami imenujemo tudi mehak prehod, za razliko od trdega prehoda, kjer se prenos za nekaj trenutkov prekine, dokler se ne vzpostavi nov kanal.

9.3.4 Želene lastnosti kod

Da bi imeli pri kodno deljenem sodostopu zgoraj omenjene prednosti, morajo imeti kode, ki jih uporabljamo za razširjanje spektra določene lastnosti. Oglejmo si zelene lastnosti kod pri

neposrednem razširjanju s kodnim zaporedjem, ki se uporablja za sodostop v tretji generaciji mobilne telefonije.

9.3.4.1 Ortogonalnost in nekoreliranost

Ker so pri kodno deljenem sodostopu uporabniki med seboj ločeni po kodah, mora imeti vsak uporabnik skupnega prenosnega medija svojo kodo. Da bi lahko pri sprejemu s pomočjo korelacije ločili uporabnike, morajo biti med seboj ortogonalne, kar pomeni, da mora biti korelacija med kodama poljubnih dveh uporabnikov enaka nič:

$$R_{ik} = \frac{1}{T_s} \int_{T_s} c_i(t)c_k(t)dt = 0 ; i \neq k \quad (9.5)$$

kjer sta $c_i(t)$ in $c_k(t)$ kodi i -tega in k -tega uporabnika, T_s pa je čas prenosa enega simbola. Vzemimo, da sprejemamo na sprejemniku signal $y(t)$ v katerem sta prisotni kodi i -tega in k -tega uporabnika:

$$y(t) = x_i[n]c_i(t) + x_k[n]c_k(t) \quad (9.6)$$

kjer sta $x_i[n]$ in $x_k[n]$ niza simbolov, ki pripadata tema dvema uporabnikoma. Pri sprejemu n -tega simbola i -tega uporabnika delamo korelacijo sprejetega signala in njegove kode čez en simbolni interval:

$$R_{yi}[n] = x_i[n]R_{ii} + x_k[n]R_{ik} \quad (9.7)$$

Če sta kodi ortogonalni, ostane v zgornjem izrazu samo prvi člen. Po preureditvi dobimo:

$$x_i[n] = \frac{R_{yi}[n]}{R_{ii}} \quad (9.8)$$

kjer je R_{ii} avtokorelacija kode enaka moči kodnega signala, ki je običajno za vse uporabnike enaka.

Ortogonalnih kod je v določenem časovnem pasu pri določeni simbolni hitrosti enako kot je neodvisnih frekvenčno ali časovno deljenih kanalov. V kodno deljenem sodostopu lahko torej z ortogonalnimi kodami ustvarimo enako število kanalov kot v časovno ali frekvenčno deljenem sodostopu. Obstaja pa lahko veliko večje število približno ortogonalnih ali semi-ortogonalnih kod, to je kod, ki imajo majhno medsebojno korelacijo. Pri uporabi približno ortogonalnih kod se uporabniki med seboj deloma motijo.

Da bi bile kode različnih uporabnikov tudi pri sprejemu ortogonalne morajo biti med seboj sinhronizirane. Če so kode med seboj časovno zamaknjene lahko namreč izgubimo ortogonalnost. Spomnimo se primera sinusa in kosinusa, ki sta med seboj ortogonalna, če pa kosinus zakasnimo za pol periode postane enak sinus in nanj ni več ortogonalen.

Da bi se izognili problemom v zvezi s sinhronizacijo morajo biti kode različnih uporabnikov med seboj nekorelirane, kar pomeni da mora biti korelacijska funkcija $R_{ik}(\tau)$ enaka nič pri vseh časovnih zamikih τ in ne samo pri zamiku $\tau = 0$, kot je to pri ortogonalnosti:

$$R_{ik}(\tau) = 0 ; i \neq k \quad (9.9)$$

Če naredimo Fourierov transform na obeh straneh zgornjega izraza dobimo:

$$C_i(f) \cdot C_k^*(f) = 0; i \neq k \quad (9.10)$$

Ker je produkt enak nič le, kadar je vsaj en člen enak nič, pomeni zgornji izraz, da morajo biti spektri različnih kod med seboj neprekrivajoči.

9.3.4.2 Avtokorelacijska funkcija kod

Če želimo pri sprejemu uporabiti RAKE sprejemnik mora biti avtokorelacijska funkcija kod čim ožja, da se pri sprejemu korelacije odbitih in direktnega signala med seboj ne prekrivajo, kot je to prikazano na sliki 8.21 v prejšnjem poglavju. Širina avtokorelacijske funkcije omejuje minimalno zakasnitev pri kateri še lahko pri sprejemu ločimo direktni in odbiti signal. Idealno bi torej bilo, da je avtokorelacijska funkcija enaka enotnemu impulzu:

$$R_{ii}(\tau) = R_{ii} \delta(\tau) \quad (9.11)$$

Če naredimo Fourierov transform zgornjega izraza, dobimo:

$$|C_i(f)|^2 = R_{ii} \quad (9.12)$$

Kar bi pomenilo, da je gostota močnostnega spektra vseh kod konstantna (neodvisna od frekvence), oziroma, da je spekter bel. Ta pogoj pa je v nasprotju s pogojem (9.10), ki zahteva, da so spektri vseh kod med seboj neprekrivajoči.

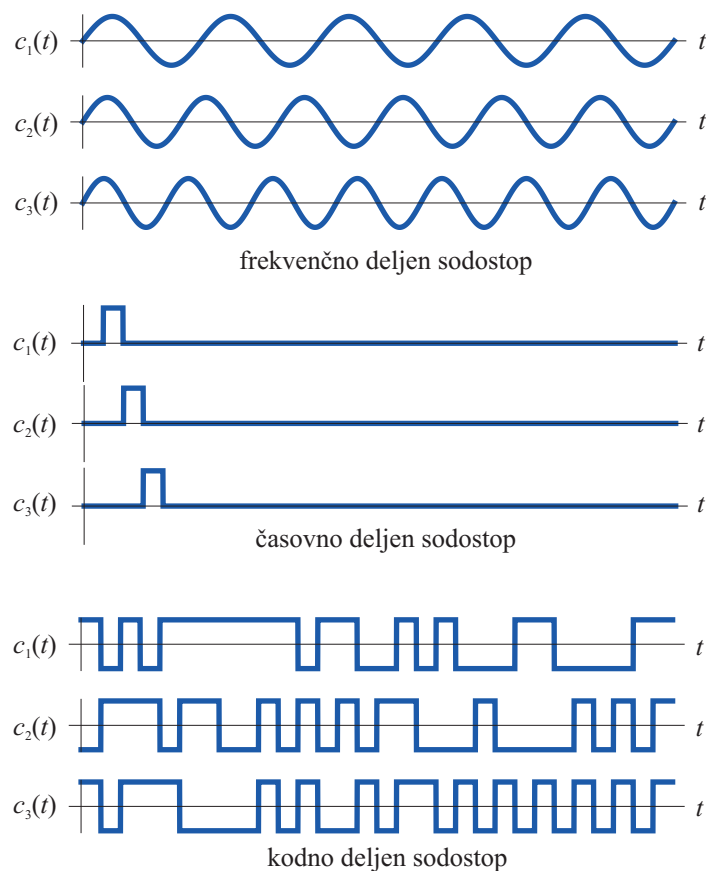
V praksi se zato uporabljajo kode, ki so približno nekorelirane in imajo približno konstanten spekter v frekvenčnem pasu namenjenem za prenos. Uporaba takih kod ima tudi prednost, da jih je precej več kot ortogonalnih, tako, da ima lahko vsak uporabnik na določeni lokaciji svojo kodo, kar omogoča ponovno uporabo istih frekvenc v sosednjih celicah in dinamično zaseganje kanala.

9.3.5 Povezava med frekvenčno, časovno in kodno deljenim sodostopom

Frekvenčno in časovno deljena sodostopa lahko jemljemo kot posebna primera kodno deljenega sodostopa z neposrednim razširjanjem s kodnim zaporedjem. Lahko rečemo, da so kode frekvenčnega sodostopa harmonični nosilci posameznih frekvenčnih kanalov, pri časovnem sodostopu pa so kode pravokotni nizi impulzov skladni s časovnimi rezinami, ki pripadajo posameznemu uporabniku. Primeri kod frekvenčno, časovno in kodno deljenega sodostopa so prikazani na sliki 9.5

Harmonični signali različnih frekvenc, ki predstavljajo kode frekvenčno deljenega sodostopa, so med seboj nekorelirani, zato pri frekvenčno deljenem sodostopu ne potrebujemo sinhronizacije. Vendar pa te kode nimajo belega spektra. Prav nasprotno. V spektru harmoničnega signala nastopa ena sama frekvenca. Zato je frekvenčno deljen sodostop zelo občutljiv na frekvenčni presih zaradi prenosa po več poteh.

Pri časovno deljenem sodostopu je spekter širok, kode so med seboj ortogonalne, niso pa nekorelirane. Če kodo enega uporabnika časovno zamaknemo, dobimo kodo drugega uporabnika. Zato je pri časovno deljenem sodostopu nujna sinhronizacija. Ker kode med seboj niso



Slika 9.5 – Primerjava kod pri frekvenčno, časovno in kodno deljenem sodostopu.

nekorelirane je časovno deljen sodostop občutljiv na prenos po več poteh. Odbiti signal lahko namreč pade v časovno okno drugega uporabnika.

Kot smo že povedali so kode pri kodnem sodostopu le približno nekorelirane in spekter je le približno konstanten. Uporabniki se zato med seboj motijo. Zaradi tega je število uporabnikov, ki hkrati dostopajo do istega prenosnega medija v določenem frekvenčnem pasu manjše kot pri časovno ali frekvenčno deljenem sodostopu, vendar lahko to razliko nadoknadimo oziroma presežemo s ponovno uporabo istih frekvenc v sosednjih celicah in z dinamičnim zaseganjem kanala.

9.4 Sodostop na osnovi prenosa z več nosilci

Sodostop na osnovi prenosa z več nosilci je posebna vrsta frekvenčno deljenega sodostopa. Pri frekvenčno deljenem sodostopu je vsakemu uporabniku dodeljen en frekvenčni kanal, medtem ko je pri sodostopu na osnovi prenosa z več nosilci vsakemu uporabniku dodeljenih več ožjih frekvenčnih podkanalov. Poleg dodeljevanja simbolov posameznim kanalom, ki je potrebno pri prenosu z več nosilci, je tu potrebno tudi dodeljevanje frekvenčnih kanalov posameznim uporabnikom.

Frekvenčne podkanale lahko dodeljujemo dinamično, glede na razmere na posameznih podkanalih, in se na ta način izogibamo frekvenčnim luknjam, ki nastanejo zaradi frekvenčnega presiha pri prenosu po več poteh. Z ustreznim dodeljevanjem frekvenčnih kanalov se lahko izogibamo tudi interferenci med sosednjimi celicami v celičnem mobilnem omrežju.

Standardizirana je digitalna izvedba sodostopa OFDMA (ang. Orthogonal Frequency Division Multiple Access). Digitalna izvedba je enaka kot pri OFDM z dodatkom dodeljevanja kanalov pri oddaji in izločanja ustreznih kanalov pri sprejemu.

OFDMA se uporablja v sistemu brezžičnega krajevnega omrežja WiMAX (Wireless Metropolitan Area Network) in v v tako imenovani 3,9 generaciji mobilne telefonije LTE (Long Term Evolution) kot tudi v nekaterih drugih manj znanih sistemih.

9.5 Naključno deljen sodostop

Naključno deljen sodostop je posebna vrsta časovno deljenega sodostopa, pri katerem vsak uporabnik zaseže celotno kapaciteto prenosnega medija za oddajo enega paketa podatkov. Naključen sodostop se uporablja predvsem v lokalnih omrežjih LAN (Local Area Network).

Glee na način, na katerega dobi posamezen uporabnik pravico, do oddaje svojega paketa, ločimo različne protokole. V nadaljevanju je podanih le nekaj najbolj znanih.

9.5.1 Protokoli ALOHA

ALOHA je prvi in najbolj preprost protokol naključno deljenega sodostopa. Razvili so ga na havajski univerzi za povezavo med havajskimi otoki. Ime protokola izvira iz havajskega pozdrava "aloha", ki pomeni nekaj podobnega kot slovenski "živijo".

Protokol deluje na osnovi oddaje s potrditvijo. Posamezen uporabnik lahko kadarkoli odda paket in čaka na potrditev sprejema. Če v istem času odda paket še kakšen drug uporabnik, pride do tako imenovanega trka, ali pa če zaradi kakega drugega razloga ne dobi potrditve v nekem vnaprej določenem času, po nekem naključno izbranem času ponovno odda paket. Čas ponovne oddaje je izbran naključno zato, da ne bi po trku oba uporabnika zopet istočasno oddala paketa in bi na ta način stalno prihajalo do ponovnih trkov. Proces se ponavlja, dokler uporabnik ne dobi potrditve.

Protokol ALOHA je primeren samo pri zelo nizki obremenitvi prenosnega kanala, saj bi pri večjem prometu stalno prihajalo do trkov, kar bi popolnoma ustavilo prenos. Prišlo bi do tako imenovanega zasičenja.

Nekoliko izboljšana verzija protokola ALOHA je razsekan ALOHA (S-ALOHA). Osnovni princip delovanja je enak kakor pri protokolu ALOHA, pri čemer pa trenutek oddaje ni več poljuben, temveč je dovoljen v točno določenih časovnih intervalih, ki so enaki dolžini paketa. Ker so paketi vseh uporabnikov časovno poravnani, lahko posamezen paket enega uporabnika pokvari samo en paket drugih uporabnikov, in ne dveh, kot je to primer, če paketi med seboj niso poravnani. To zmanjša verjetnost trka na polovico.

Razsekani ALOHA nekoliko bolje izkorišča zmogljivost prenosnega kanala, vendar pa morajo biti zato uporabniki med seboj časovno sinhronizirani, podobno kot pri časovno deljenem sodostopu.

Dodatna izboljšava protokola ALOHA je rezervacijski ALOHA (R-ALOHA). Pri tem protokolu se uporabljajo kratka časovna okna za rezervacijo daljših časovnih oken za prenos podatkov. Uporabnik najprej preko rezervacijskih oken rezervira podatkovna okna za prenos. Zato pri prenosu podatkov ne prihaja do trkov. Do trkov lahko prihaja zgolj pri kratkih rezervacijskih paketih, ki se izvajajo po protokolu S-ALOHA. Pri prenosu daljših podatkovnih paketov, ki se prenašajo v rezerviranih časovnih oknih, pa do trkov ne more priti. Zato je verjetnost trkov pri R-ALOHA bistveno manjša kot pri S-ALOHA, potrebna pa je poleg sinhronizacije še dodatna režija za rezervacijo podatkovnih oken.

9.5.2 Protokoli z zaznavanjem nosilca

Pri protokolu sodostopa z zaznavanjem nosilca CSMA (Carrier Sense Multiple Access), se skušamo izogniti trkom na ta način, da vsak uporabnik posluša kaj se dogaja na kanalu, tako da lahko zazna, ali je kanal zaseden, oziroma ali je prisoten nosilec, kadar gre za prenos v višjih frekvenčnih legah. Uporabnik odda svoj paket šele, ko je kanal prost.

Na ta način se verjetnost trkov močno zmanjša, vendar lahko do trkov še vedno prihaja zaradi zakasnitve signala na prenosni poti. To se lahko zgodi, ko en uporabnik odda svoj paket, vendar ta zaradi zakasnitve še ne prispe do drugega uporabnika, ki želi oddati svoj paket. Ker ne zazna, da je kanal zaseden, tudi on odda svoj paket in pride do trka. Vendar so ti trki precej manj verjetni kot pri protokolih ALOHA.

Pri takem načinu obstaja nevarnost, da bi več uporabnikov poslušalo kanal in čakalo na oddajo svojega paketa. Ko bi se kanal sprostil, bi vsi hkrati oddali svoje pakete in bi nujno prišlo do trka. Zato je v protokolu predvideno, da uporabnik, ko zazna, da je kanal prost, počaka še nek naključen čas in šele potem odda svoj paket.

Če po določenem času po oddaji paketa uporabnik ne dobi potrditve, ponovi celoten postopek. Ker je, ne glede na to, da je prišlo do trka, oddal cel paket, nato pa mora isti paket ponovno prenašati, se pri trku izgubi ves čas trajanja enega paketa. Izgubljeni čas se lahko bistveno skrajša pri protokolu z zaznavanjem nosilca in detekcijo trka CSMA/CD (Carrier Sense Multiple Access with Collision Detection). Tu uporabnik tudi sam sprejema svoj oddani paket. Ko ugotovi, da je paket zaradi trka pokvarjen, takoj prekine prenos in počaka nek naključen čas pred ponovno oddajo paketa. Ker pride do trkov lahko samo zaradi zakasnitve na prenosni poti, čas od začetka prenosa do detekcije trka ni daljši od zakasnitve, ta pa je običajno precej manjša od dolžine paketa.

Ker se verjetnost trkov povečuje z zakasnitvijo na prenosni poti, ta pa se povečuje z razdaljo med uporabniki, je protokol CSMA/CD primeren le za relativno majhne razdalje v lokalnih omrežjih. Uporabljen je na primer v lokalnem omrežju Ethernet.

Največja pomanjkljivost protokola CSMA/CD je v tem, da lahko pri velikem prometu do zasičenja omrežja. Na omrežju stalno prihaja do trkov in prenos se popolnoma ustavi. Zato

je možno prenosni kanal izkoristiti le do približno 75% njegove zmogljivosti. V praksi delujejo lokalna omrežja pri precej nižjih obremenitvah, tako da to ni običajno problem. Problem v zvezi s tem pa lahko nastopi ob posebnih situacijah, ko se promet nenadoma, zaradi nekih zunanjih razlogov, kot je na primer neka naravna nesreča, močno poveča.

9.5.3 Zaseganje na osnovi žetona

Protokol z zaseganjem na osnovi žetona (ang. token) preprečuje možnost trkov v omrežju. Virtualen žeton tu kroži od uporabnika do uporabnika po oštevilčenem vrstnem redu. Samo uporabnik, ki ima v danem trenutku žeton, sme oddati svoj paket podatkov. Če ima paket pripravljen za oddajo, ga odda, drugače pa preda žeton naslednjemu uporabniku.

Ker ima žeton lahko samo en uporabnik, ne more priti do trkov. Kadar je promet relativno majhen, veliko uporabnikov ne odda paketa, temveč samo preda žeton naslednjemu uporabniku in se postavi na konec vrste uporabnikov, ki čakajo na žeton. Uporabnik, ki ima potrebo po oddaji več paketov, mora po vsakem paketu počakati, da pride žeton zopet do njega. Pri nizkih obremenitvah omrežja so zato zakasnitve večje kot pri prej opisanih protokolih.

Ko promet narašča, pa pri zaseganju na osnovi žetona ne pride do zasičenja omrežja, temveč se samo podaljšujejo zakasnitve. Protokol na osnovi zaseganja z žetonom tudi omogoča, da se v nujnih primerih vzpostavi prioriteta lista in prejmejo žeton samo uporabniki z visoko prioriteto. Zato se protokol na osnovi zaseganja z žetonom ponavadi uporablja v omrežjih, kjer bi bilo zasičenje omrežja v posebnih situacijah lahko kritično, kot so to na primer omrežja na letališčih, bankah, urgentnih bolnišnicah in podobno. Primer omrežja na tej osnovi je IBM Token Ring, ki se večinoma uporablja v bankah.

Teorija informacij

V poglavju je opredeljen pojem informacije. Podana je matematična definicija informacije na osnovi verjetnosti. Opredeljen je izvor informacije in entropija izvora informacije kot povprečna informacija simbolov na izhodu izvora informacij. Na osnovi pogojne verjetnosti je definirana vzajemna informacija in vzajemna entropija digitalnega kanala. Podane so definicija kapacitete digitalnega kanala in kapaciteta frekvenčno omejenega prenosnega kanala. Definiran je Gaussov kanal in podane enačbe za njegovo kapaciteto in spektralno učinkovitost.

Kot smo omenili že na začetku, pri obravnavi osnovnega telekomunikacijskega modela, gre v telekomunikacijah za prenos sporočila od njegovega izvora do prejemnika sporočila. Da je neko sporočilo smiselno prenašati, mora to vsebovati nekaj, kar prejemniku sporočila ni znano oziroma tega ne bi mogel ugotoviti, če ne bi sprejel sporočila. Imenujmo to informacija. Informacija, ki jo dobi prejemnik sporočila je torej odvisna tudi od tega, kaj prejemnik o sporočilo že ve oziroma kaj pričakuje, kar pa je, če je prejemnik človek, zelo subjektivno.

V tem poglavju bomo opredelili informacijo v bolj tehničnem smislu. Predpostavili bomo, da je prejemnik informacije neka naprava, ki ima lahko neko vnaprejšnje, statistično vedenje o vrsti sporočil, ki se pošiljajo, ne bomo pa se ukvarjali z dejansko vsebino in pomenom sporočila. Vnaprejšnje vedenje o vrsti sporočila je namreč mogoče izkoristiti za izboljšanje učinkovitosti in/ali zanesljivosti prenosa, kar lahko dosežemo z ustreznimi kodirnimi postopki, ki so obravnavani v naslednjem poglavju.

10.1 Verjetnost in informacija

Kot smo že omenili, je informacija v sporočilu tisto, česar prejemnik sporočila ne ve vnaprej. V zvezi s tem pa se postavi pa se vprašanje, kolikšna je ta informacija oziroma kako kvantitativno opredeliti informacijo. Informacijo bomo zato opredelili kot merilo presenečenja. Če prejemnik natanko ve, kakšno bo sporočilo, ki ga bo prejel, potem v sporočilu ni nič novega in ni dobil nobene informacije. Kolikor manj ko pričakuje določeno sporočilo, toliko bolj je presenečen in toliko večja je informacija, ki jo je dobil.

Presenečenje je povezano z verjetnostjo. Bolj ko je določeno sporočilo verjetno, bolj k ga pričakujemo, manjše je presenečenje, ko ga dobimo. Da bi lahko opredelili informacijo, moramo zato najprej opredeliti verjetnost.

10.2 Verjetnost

V tehniki se pojem verjetnosti enačim s pojmom pogostosti. Za dogodke (sporočila ki jih sprejmemo), ki se zgodijo pogosto pravimo da so bolj verjetni, dogodki, ki se zgodijo redko, pa so manj verjetni.

Verjetnost kot pogostost je definirana s statistično definicijo verjetnosti. Statistična definicija verjetnosti temelji na velikem številu ponovitev istega poskusa. Rezultat posameznega poskusa imenujemo dogodek. Število različnih možnih dogodkov poskusa mora biti omejeno, kot na primer pri metu kovanca ali kocke. Pri metu kovanca lahko pade grba ali glava, pri metu kocke pa so to lahko številke ena do šest. Pri prenosu lahko obravnavamo kot dogodek vrednost posameznega simbola v sporočilu ali pa tudi vrednost celotnega sporočila. Ker je celotno sporočilo sestavljeno iz več simbolov (na primer črk v tekstu) pravimo, da je tak dogodek sestavljen dogodek, to je dogodek, ki vsebuje več osnovnih dogodkov, to je simbolov sporočila.

Ker govorimo o prenosu sporočil, bomo definirali verjetnost kar na osnovi simbolov. Simbol lahko zavzame le eno od vnaprej določenih vrednosti. Nabor vseh možnih vrednosti imenujemo abeceda. Abeceda dvonivojskih simbolov je na primer lahko $\{0,1\}$, abeceda štirinivojskih simbolov pa $\{-3, -1, 1,3\}$. V splošnem lahko posamezen simbol, označimo ga z x , zavzame eno od L vrednosti abecede:

$$x \in \{x_1, x_2, x_3, \dots, x_L\} \quad (10.1)$$

kjer so z x_1, x_2, \dots označene možne vrednosti simbola.

Kakšno vrednost bodo imeli posamezni simboli je odvisno od izvora sporočila. Izvor sporočila je lahko neka knjiga, datoteka na računalniku ali pa preprosto metanje kovanca ali kocke. Če bi prenašali rezultate metanja kovanca, bi lahko na primer prenesi simbol 0, kadar pade glava in simbol 1 kadar pade grb. Ne glede na to, kaj je izvor sporočila, pa je za prejemnika informacije vrednost posameznih simbolov naključna. Statistično definiramo verjetnost, da bo simbol x zavzel vrednost x_k iz abecede, kot:

$$P_x[k] = \lim_{N \rightarrow \infty} \frac{N_k}{N} \quad (10.2)$$

kjer je N število vseh opazovanih simbolov, in N_k število simbolov, ki so imeli vrednost x_k . Definicija sicer zahteva, da gre N proti neskončno, kar pa v praksi seveda ni mogoče. V praksi želimo da je N čim večji, da se čim bolj približamo dejanski vrednosti verjetnosti.

Če se nek dogodek vedno zgodi (gotov dogodek) potem je N_k enak N in verjetnost gotovega dogodka je enaka 1. Če se dogodek nikoli ne zgodi (nemogoč dogodek) potem je N_k enak nič in tudi verjetnost nemogočega dogodka je enaka 0. Vsi ostali dogodki imajo verjetnost med temi dvema skrajnimi vrednostmi.

Kadar nimamo na voljo sporočil, na osnovi katerih bi lahko določili verjetnost (o sporočilu ne vemo popolnoma nič), predpostavimo, da simboli enako pogosto zavzamejo vsako vrednost iz abecede. Potem velja:

$$N_k = \frac{N}{L} \quad (10.3)$$

in

$$P_x[k] = \frac{1}{L} \quad (10.4)$$

Verjetnost $P_x[k]$ je funkcija indeksa k in jo imenujemo tudi porazdelitev verjetnosti. Kadar so vse vrednosti simbola enako verjetne, kot je to v zgornjem primeru, pravimo, da je porazdelitev verjetnosti enakomerna.

10.3 Odvisni simboli in pogojna verjetnost

Posamezni simboli so lahko med seboj odvisni. V tekstu, na primer, lahko nastopajo samo določene kombinacije črk. V slovenskem jeziku se na primer črke zelo redko ponavljajo. Če je določena črka v tekstu A je verjetnost, da bo naslednja črka tudi A ali nek drug samoglasnik bistveno manjša kot, da bo naslednja črka soglasnik. Če pa simboli predstavljajo met kovanca ali kocke, pa lahko rečemo, da so med seboj neodvisni, saj rezultat prejšnjega meta nikakor ne vpliva na rezultat naslednjega meta.

Med seboj so lahko odvisni simboli v istem nizu, kot je to primer za črke v knjigi, kakor tudi simbol različnih nizov, kot sta to na primer niza simbolov na vhodu in izhodu digitalnega prenosnega kanala. Verjetnost, da bo nek simbol y zavzel vrednost y_m , če vemo, da je simbol x zavzel vrednost x_k , imenujemo pogojna verjetnost, in jo označimo z $P_{y|x}[m, k]$. Če opazujemo oba simbola kot en dogodek, dobimo sestavljen dogodek $\{x, y\}$. Rezultat tega dogodka je vrednost obeh simbolov. Verjetnost sestavljenega dogodka, to je verjetnost da bo imel x vrednost x_k in hkrati y vrednost y_m je potem enaka:

$$P_{x,y}[k, m] = P_x[k] P_{y|x}[m, k] \quad (10.5)$$

Zgornjo enačbo imenujemo tudi Bayasevo pravilo. Če sta simbola med seboj neodvisna, potem je verjetnost da je simbol y zavzel vrednost y_m neodvisna od vrednosti, ki jo je zavzel simbol x , torej za neodvisna simbola velja:

$$P_{y|x}[m, k] = P_y[m] \quad (10.6)$$

in

$$P_{x,y}[k, m] = P_x[k] P_y[m] \quad (10.7)$$

Verjetnost sestavljenega dogodka je v tem primeru kar enaka produktu posameznih dogodkov. Pri metu kocke je na primer verjetnost, da bo padla šest enaka $1/6$. Ker sta dva meta med seboj neodvisna, je verjetnost, da bo dvakrat padla šest torej enaka $1/36$.

10.4 Srednja vrednost, srednja kvadratična vrednost in varianca

Če poznamo porazdelitev verjetnosti simbolov v nizu, to je pogostost s katero nastopajo posamezne vrednosti, lahko določimo tudi srednjo vrednost posameznega simbola po enačbi:

$$\bar{x} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N x_n = \sum_{k=1}^L x_k P_x[k] \quad (10.8)$$

Na podoben način lahko določimo srednjo kvadratično vrednost:

$$\overline{x^2} = \sum_{k=1}^L x_k^2 P_x[k] \quad (10.9)$$

ki ustreza moči digitalnega signala in varianco:

$$\sigma_x^2 = \text{var}(x) = \sum_{k=1}^L (x_k - \bar{x})^2 P_x[k] \quad (10.10)$$

Varianca predstavlja moč digitalnega signala, ki smo mu odšteli srednjo vrednost oziroma moč njegovega odstopanja od srednje vrednosti. S σ_x smo označili standardno odstopanje oziroma standardno deviacijo. Standardna deviacija je enaka kvadratnemu korenu variance. Standardna deviacija se pogosto uporablja v statistiki kot merilo, kako dobro srednja vrednost predstavlja vse posamezne vrednosti. Če so odstopanja od srednje vrednosti velika, potem je velika tudi standardna deviacija.

10.5 Informacija

Kot smo že omenili je informacija merilo presenečenja. Bolj verjetni dogodki imajo zato manjšo informacijo, manj verjetni dogodki pa večjo. Definirajmo informacijo simbola x z vrednostjo x_k kot:

$$I_x[k] = \log_2 \frac{1}{P_x[k]} \quad (10.11)$$

Oglejmo si nekaj lastnosti, ki jih ima informacija po tej definiciji.

- Ker je verjetnost gotovega dogodka enaka 1, je njegova informacija po zgornji enačbi enaka 0. Če je dogodek gotov, namreč vrednost simbola vnaprej poznamo, torej nismo nič presenečeni, ko jo izvemo.
- Ker nastopa verjetnost v imenovalcu zgornjega izraza, je informacija večja, kadar je verjetnost manjša. Manj verjetni dogodki so bolj presenetljivi.
- Ker je verjetnost število ki med 0 in 1, je informacija vedno pozitivna, razen pri gotovem dogodku, ko je enaka 0.
- Kadar imamo dva neodvisna simbola, je verjetnost sestavljenega dogodka enak produktu posameznih verjetnosti. Ker je logaritem produkta enak vsoti logaritmov posameznih členov, je informacija sestavljenega dogodka enaka vsoti informacij posameznih simbolov.
- Kadar ima simbol samo dve možni vrednosti in sta obe enako verjetni, je njuna verjetnost enaka 1/2, njuna informacija pa je enaka 1. Enota za informacijo je bit ali tudi Shannon (sh). En bit informacije dobimo torej, ko zremo odgovor na vprašanje, kjer sta odgovora da in ne enako verjetna.
- Ker pomeni sprememba osnove logaritma samo množenje s konstanto, lahko informacijo definiramo z logaritmom s poljubno osnovo. S spremembo osnove se spremeni samo enota, s katero merimo informacijo. Za naravni logaritem je enota nat.

10.6 Entropija

V digitalnem signalu oziroma nizu simbolov nastopajo simboli z vrednostmi iz ustrezne abecede. Pogostost, s katero nastopajo posamezne vrednosti iz abecede, je določena s porazdelitvijo verjetnosti. Informacija simbola, ki ima določeno vrednost, je odvisna od verjetnosti, s katero ta vrednost nastopa, zato v splošnem simboli v nizu simbolov nimajo enake informacije. Ker je vrednost posameznega simbola naključna, je tudi njegova informacija naključna. Za posamezen simbol v nizu torej ne moremo vnaprej vedeti, kolikšna bo njegova informacija.

Vendar, če poznamo porazdelitev verjetnosti, vemo kako pogosto nastopajo posamezne vrednosti in s tem tudi kako pogosto imajo simboli informacijo, ki ustreza tem vrednostim. Za niz med seboj neodvisnih simbolov lahko izračunamo povprečno informacijo simbola v nizu. Povprečna informacija imenujemo entropija in je dana z izrazom:

$$H_x = \sum_{k=1}^L P_x[k] I_x[k] = \sum_{k=1}^L P_x[k] \log_2 \frac{1}{P_x[k]} \quad (10.12)$$

Pri enakomerni porazdelitvi verjetnosti, so vse vrednosti simbolov enako verjetne z verjetnostjo $1/L$ zato je entropija pri enakomerni porazdelitvi enaka:

$$H_x = \sum_{k=1}^L \frac{1}{L} \log_2 L = \log_2 L \quad (10.13)$$

Pokazati je mogoče, da je to tudi maksimalna entropija, ki jo ima lahko niz simbolov z dolžino abecede enako L .

10.7 Redundanca in irelevanca

Kadar porazdelitev ni enakomerna oziroma, če so simboli med seboj niso neodvisni, potem je entropija niza manjša od maksimalne entropije $1/L$. Informacija celotnega niza je zato manjša, kot bi lahko bila. Tak niz bi lahko nadomestili z nizom z manj simboli in/ali z nizom s krajšo abecedo. Odvečnost v zapisu informacije imenujemo redundanca. Redundanca je definirana kot:

$$R_x = 1 - \frac{H_x}{H_{max}} = 1 - H_x L \quad (10.14)$$

Kadar je entropija niza enaka maksimalni entropiji je redundanca enaka 0 (ni redundance), kadar pa je entropija niza enaka 0, je redundanca enaka 1 (100% redundanca).

Z ustreznimi kodirnimi postopki je mogoče kodirati niz tako, da izločimo redundanco in zato za zapis potrebujemo manj simbolov. Tako kodiranje imenujemo tudi entropijsko kodiranje. V osnovnem telekomunikacijskem modelu je za to zadolžen kodirnik izvora.

Ker v okviru te knjige izvirnega kodiranja ne bomo natančneje obravnavali, omenimo le, da programi za stiskanje datotek, kot sta to kot so zip ali rar, izločajo redundanco in na ta način skrčijo datoteke na računalniku. Tako krčenje pride prav pri prenosu podatkov, saj potrebno prenesti manjše število simbolov in zato potrebujemo manjše prenosne kapacitete.

Pri prenosu lahko izvorni kodirnik poleg redundance izloča tudi irelevanco. Irelevanca imenujemo tisto, kar je za prejemnika nepomembno. Če je to zvok, je nepomembno tisto, česar prejemnik ne more slišati, zaradi omejenih sposobnosti slušnih organov, če gre za prenos slike, pa je nepomembno tisto, česar prejemnik ne more videti zaradi omejenih sposobnosti človeka za dojetanje slik. Na tej osnovi delujejo kodirniki zvoka, kot to mp3, in kodirniki slike, kot je to na primer mp4. Izločanje irelevance lahko še bistveno bolj skrči podatke kot izločanje redundance, vendar lahko pri zelo velikem krčenju kvaliteta zvoka ali slike nekoliko pade. To se zgodi, kadar smo zaradi močnem krčenju izločili tudi del koristne informacije.

10.8 Vzajemna informacija

Kadar sta simbola x in y med seboj odvisna, imamo določeno informacijo o simbolu x , če poznamo vrednost simbola y in obratno. To informacijo imenujemo vzajemna informacija. Informacija, ki jo ima simbol y z vrednostjo y_m o simbolu x z vrednostjo x_k zapišemo v obliki $I_{y|x}[m, k]$.

Omenili smo že, da je skupna informacija dveh simbolov enaka vsoti posameznih informacij, če sta simbola neodvisna.

$$I_{x,y}[k, m] = I_x[k] + I_y[m] \quad (10.15)$$

Kadar sta simbola med seboj odvisna, pa je njuna skupna informacija nekoliko manjša. Vzajemna informacija, to je informacija, ki jo ima y o x , je vsebovana že v informaciji x -a, zato jo moramo od informacije y -a odšteti, preden to prištejemo skupni informaciji. Dobimo:

$$I_{x,y}[k, m] = I_x[k] + I_y[m] - I_{y|x}[m, k] \quad (10.16)$$

Vidimo, da je skupna informacija dveh simbolov manjša, kadar sta simbola med seboj odvisna. To je tudi razlog, da imajo nizi odvisnih simbolov manjšo entropijo kot nizi neodvisnih simbolov.

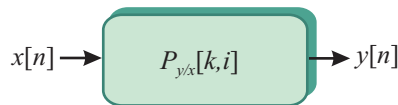
Iz zgornjega izraza lahko sedaj izrazimo vzajemno informacijo:

$$I_{y|x}[m, k] = I_x[k] + I_y[m] - I_{x,y}[k, m] \quad (10.17)$$

Z upoštevanjem definicije informacije in Bayasovega pravila:

$$I_{y|x}[m, k] = \log_2 \frac{P_{x,y}[k, m]}{P_x[k] P_y[m]} = \log_2 \frac{P_{y|x}[k, m]}{P_y[m]} \quad (10.18)$$

Za razliko od informacije, za katero smo povedali, da je vedno pozitivna ali enaka 0, pa je lahko vzajemna informacija tudi negativna. Negativna vzajemna informacija predstavlja dezinformacijo. Ko smo zvedeli vrednost simbola y , ki nastopa z zelo majhno verjetnostjo pri dejanski vrednosti simbola x , smo dobili o simbolu x napačno informacijo, saj sklepamo, da je zelo malo verjetno, da ima vrednost, ki pa jo v resnici ima.



Slika 10.1 – Digitalni kanal preslika niz simbolov na svojem vhodu v niz simbolov na svojem izhodu

10.9 Digitalni kanal

Digitalni kanal je kanal, ki preslika niz simbolov na svojem vhodu v niz simbolov na svojem izhodu, kot je to prikazano na sliki 10.1.

Celoten sistem na sliki 7.4, ki smo ga obravnavali pri prenosu v osnovnem pasu je primer digitalnega kanala. Kadar je tak prenos brez intersimbolne interference, je simbol $y[n]$ na izhodu kanala odvisen samo od pripadajočega simbola $x[n]$ na izhodu kanala. Pravimo, da je kanal brez spomina, ker si ne zapomni preteklih vrednosti simbolov, tako da te ne morejo vplivati vplivati na trenutni simbol. Tak kanal imenujemo zato digitalni kanal brez intersimbolne interference ali digitalni kanal brez spomina.

Čeprav je trenutni simbol na izhodu odvisen samo od trenutnega simbola na izhodu, lahko pride zaradi šuma in motenj na kanalu z določeno verjetnostjo do napake pri prenosu. Digitalni kanal brez spomina lahko zato opišemo s porazdelitvijo pogojnih verjetnosti $P_{y|x}[m, k]$, ki pove s kakšno verjetnostjo bo imel simbol na izhodu vrednost y_m , če je vrednost simbola na vhodu x_k , za vse možne pare m in k . Rečemo lahko, da porazdelitev $P_{y|x}[m, k]$ torej popolnoma opisuje tak kanal, na podoben način, kot smo rekli, da sistemska funkcija popoloma opisuje linearen in časovno neodvisen kanal.

Simboli na izhodu so odvisni od simbolov na vhodu. Vzemna informacija je informacija, ki jo imajo simboli na izhodu o simbolih na vhodu. Vzemna informacija je torej informacija, ki smo jo prenašamo preko digitalnega kanala.

10.10 Vzajemna entropija

Koliko informacije prenesemo s prenosom posameznega simbola je odvisno od dejanskih vrednosti vhodnega in izhodnega simbola. Ker so simboli naključni, te vrednosti vnaprej ne moremo določiti, lahko pa izračunamo povprečno vzajemno informacijo oziroma vzajemno entropijo:

$$H_{y|x} = \sum_{k=1}^L \left[P_x[k] \sum_{m=1}^M P_{y|x}[m, k] \log_2 \frac{P_{y|x}[m, k]}{P_y[m]} \right] \quad (10.19)$$

Kjer sta z L in K označeni dolžini abeced vhodnih in izhodnih simbolov, ki nista nujno enaki.

Simboli na izhodu kanala so odvisni od vhodnih simbolov in od kanala, zato lahko porazdelitev $P_y[m]$ izrazimo s porazdelitvima $P_x[k]$ in $P_{y|x}[m, k]$ kot:

$$P_y[m] = \sum_{k=1}^L P_{y|x}[m, k] P_x[k] \quad (10.20)$$

Če sedaj Vstavimo zgornji izraz v enačbo (10.21), dobimo:

$$H_{y|x} = \sum_{k=1}^L P_x[k] \sum_{m=1}^M \left(P_{y|x}[m, k] \log_2 \frac{P_{y|x}[m, k]}{\sum_{k=1}^L P_{y|x}[m, k] P_x[k]} \right) \quad (10.21)$$

Zgornji izraz omogoča izračunati povprečno informacijo, ki jo prenesemo z enim simbolom preko kanala, za katerega poznamo porazdelitev verjetnosti $P_{y|x}[m, k]$.

10.11 Kapaciteta digitalnega kanala

V prejšnjem razdelku smo videli, da je vzajemna entropija $H_{y|x}$ digitalnega kanala, to je informacija, ki jo v povprečju prenesemo s prenosom enega simbola, odvisna tako od samega kanala, ki ga opisuje porazdelitev pogojne verjetnosti $P_{y|x}[m, k]$, kakor tudi od porazdelitve verjetnosti simbolov na vhodu $P_x[k]$, to je od tega, kar smo pripeljali na vhod kanala.

Predpostavimo sedaj, da imamo na voljo digitalni kanal, ki ga ne moremo spreminjati, želeli pa bi ga čim boljše izkoristiti. To pomeni, da bi želeli preko tega kanala prenesti čim več informacije, ali z drugimi besedami, maksimirati vzajemno entropijo.

Ker je vzajemna entropija odvisna odvisna tudi od porazdelitve verjetnosti na vhodu, obstaja neka optimalna porazdelitev $P_x[k]_{opt}$, pri kateri je vzajemna entropija danega kanala maksimalna. Vzajemno entropijo, ki jo dobimo pri optimalni porazdelitvi na vhodu imenujemo kapaciteta digitalnega kanala C_d , kar lahko zapišemo kot

$$C_d = \max_{P_x[k]} H_{y|x} \quad (10.22)$$

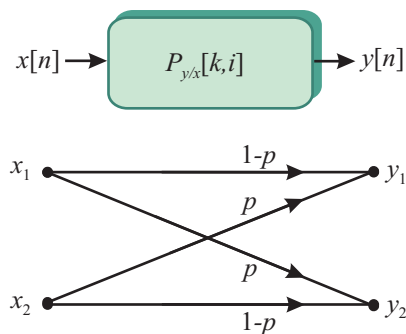
Kapaciteta digitalnega kanala je maksimalna vzajemna entropija tega kanala, pri čemer gre maksimizacija preko vseh možnih porazdelitev verjetnosti vhodnih simbolov. Kapaciteta kanala pove kolikšna je največja možna povprečna informacija, ki jo lahko prenesemo z enim simbolom preko takega kanal. Tej meji se je mogoče približati z ustreznimi postopki kanalskega kodiranja, ni pa je mogoče preseči.

Pri digitalnem kanalu imamo omejen nabor vrednosti vhodnih in izhodnih simbolov. Količina informacije, ki jo lahko prenesemo preko takega kanala je odvisna tudi od tega, kakšna sta ta dva nabora vrednosti. Z binarnimi simbolom namreč ne moremo prenašati več kot enega bita informacije, saj je entropija binarnih simbolov največ 1.

Če sprostimo omejitve, ki določa vrednosti simbolov na vhodu in izhodu kanala in dovolimo, da so to poljubna realna števila, lahko povečamo kapaciteto kanala. Kanal ni več digitalen ampak samo časovno diskreten. Kapaciteta časovno diskretnega kanala bo večja ali kvečjemu enaka kot kapaciteta digitalnega kanala, saj vsak digitalni kanal predstavlja samo poseben primer časovno diskretnega kanala. Z digitalnim kanalom se lahko poljubno približamo časovno diskretnemu kanalu z dovolj velikima abecedama vhodnih in izhodnih simbolov.

10.12 Binarni simetrični kanal

Preprost primer digitalnega prenosnega kanala, ki pa v praksi zelo pogosto nastopa, je binarni simetrični kanal (BSC - Binary Symmetric Channel), ki je prikazan na sliki 10.2.



Slika 10.2 – Binarni simetrični kanal. Verjetnost napake p , je verjetnost, da bo pri oddanem x_1 sprejet y_2 oziroma pri oddanem x_2 sprejet y_1 .

Pri binarnem simetričnem kanalu sta vhodni in izhodni niz binarna, torej imata vsak po dve možni vrednosti $\{x_1, x_2\}$ in $\{y_1, y_2\}$.

Verjetnost, da simbol na prehodu skozi kanal spremeni vrednost, je enaka p in verjetnost, da ostane vrednost nespremenjena, je enaka $1 - p$. Porazdelitev pogojne verjetnosti je potem podana z:

$$\begin{aligned}
 P_{y|x}[1, 1] &= 1 - p \\
 P_{y|x}[1, 2] &= p \\
 P_{y|x}[2, 1] &= p \\
 P_{y|x}[2, 2] &= 1 - p
 \end{aligned} \tag{10.23}$$

Označimo še s q verjetnost, da ima vhodni simbol vrednost x_1 . Porazdelitev verjetnosti na vходу je potem enaka:

$$\begin{aligned}
 P_x[1] &= q \\
 P_x[2] &= 1 - q
 \end{aligned} \tag{10.24}$$

Ko vstavimo (10.23) in (10.24) v (10.21) dobimo za vzajemno entropijo binarnega simetričnega kanala izraz:

$$\begin{aligned}
 H_{y|x}(p, q) &= q(1-p) \log_2 \frac{1-p}{q+p-2pq} + \\
 &+ qp \log_2 \frac{p}{1-q-p+2pq} + \\
 &+ (1-q)p \log_2 \frac{p}{q+p-2pq} + \\
 &+ (1-q)(1-p) \log_2 \frac{1-p}{1-q-p+2pq}
 \end{aligned} \tag{10.25}$$

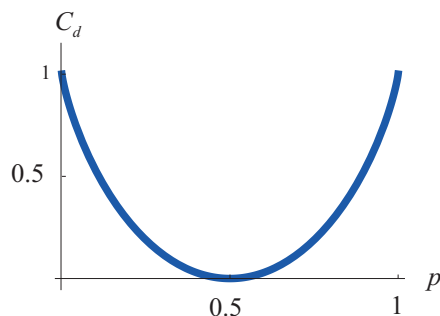
Da bi lahko izračunali kapaciteto binarnega simetričnega kanala, moramo določiti q tako, da bo vrednost zgornjega izraza maksimalna. Vrednost q_{opt} , pri katerem je vzajemna entropija maksimalna, dobimo tako, da zgornji izraz odvajamo po q in odvod izenačimo z 0. Dobimo:

$$q_{opt} = \frac{1}{2} \quad (10.26)$$

Kapaciteto binarnega simetričnega kanala dobimo, če vstavimo optimalno vrednost q_{opt} v izraz za vzajemno entropijo (10.25):

$$C_d = (1 - p) \log_2 2(1 - p) + p \log_2 2p \quad (10.27)$$

Kapaciteta binarnega simetričnega kanala kot funkcija verjetnosti p je prikazana na sliki ???. Vidimo, da je kapaciteta maksimalna pri $p = 0$ in $p = 1$, kjer je enaka 1 najmanjša pa je pri $p = 1/2$, kjer je enaka 0. Na prvi pogled se zdi morda nenavadno, da je kapaciteta kanala enaka



Slika 10.3 – Kapaciteta binarnega simetričnega kanala. Kapaciteta je odvisna od samega kanala, to je od verjetnosti p , da pride do napake. Kapaciteta je enaka 1 bit/simbol pri $P = 0$ in $p = 1$ in enaka 0 pri $p = 1/2$.

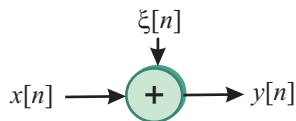
pri verjetnosti napake $p = 0$, torej takrat, ko pri prenosu ne pride do napak, in pri verjetnosti napake $p = 1$, to je takrat, ko vedno pride do napake. Vendar, kadar že vnaprej vemo, da bo pri prenosu x_1 na izhodu y_2 in obratno, lahko iz sprejetih simbolov natančno ugotovimo, kateri simboli so bili oddani. To pomeni, da smo dobili enako informacije kot v primeru prenosa brez napake.

10.12.1 Gaussov kanal

Gaussov kanal je kanal brez spomina. Simboli na vhodu in izhodu nimajo omejene abecede, zato je to diskreten kanal. Vhodnim $x[n]$ simbolom se na prenosni poti prištejejo vzorci belega Gaussovega šuma (AWGN - Additive White Gaussian Noise) $\xi[n]$. Vzorci so med seboj neodvisne zvezne naključne spremenljivke z Gaussovo porazdelitvijo gostote verjetnosti. Več o zvenih naključnih spremenljivkah in Gaussovi porazdelitvi si lahko bralec prebere v kateri koli knjigi o verjetnosti. Tu omenimo le, da ima termični šum in tudi šum, ki nastopa kot posledica več med seboj neodvisnih virov Gaussovo ali približno Gaussovo porazdelitev verjetnosti. Shematsko je Gaussov kanal prikazan na sliki 10.4

Gaussov kanal opisuje enačba:

$$y[n] = x[n] + \xi[n] \quad (10.28)$$



Slika 10.4 – Gaussov kanal. Simbolom $x[n]$ se na prenosni poti prištejejo vzorci Gaussovega šum $\xi[n]$.

Za diskretni kanal lahko izračunamo vzajemno entropijo tako, da izračunamo limito izraza za vzajemno entropijo digitalnega kanala ko gresta L in M proti neskončno. Vsote pri tem preidejo v integrale, porazdelitve verjetnosti pa v gostote porazdelitev verjetnosti. Izpeljava za diskretni kanal presega okvire te knjige.

Kapaciteto diskretnega kanala je v splošnem izredno težko določiti, saj je nabor možnih porazdelitev vhodnih simbolov neomejen, zato je le v posebnih primerih možno analitično poiskati optimalno porazdelitev na vходу. Gaussov kanal je eden od posebnih primerov, za katerega je možno analitično izračunati kapaciteto. Po daljši izpeljavi dobimo za kapaciteto Gaussovega kanala naslednji izraz:

$$Cd = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_x^2}{\sigma_\xi^2} \right) \quad (10.29)$$

kjer je σ_x^2 varianca oziroma moč signala in σ_ξ^2 varianca oziroma moč šuma. Razmerje σ_x^2/σ_ξ^2 imenujemo razmerje med signalom in šumom in označujemo z SNR (Signal to Noise Ratio). Vidimo, da je kapaciteta Gaussovega signala neposredno odvisna od razmerja med signalom in šumom:

$$Cd = \frac{1}{2} \log_2 (1 + SNR) \quad (10.30)$$

10.13 Kapaciteta frekvenčno omejenega analognega Gaussovega kanala

Pri frekvenčno omejenem analognem kanalu frekvenčna širina omejuje simbolno hitrost pri prenosu brez intersimbolne interference. V skladu z Nyquistovim kriterijem je maksimalna simbolna hitrost enaka:

$$f_{smax} = 2B \quad (10.31)$$

kjer je B pasovna širina kanala. Maksimalna hitrost, s katero lahko prenašamo informacijo preko takega kanala imenujemo kapaciteta kanala C . Ta je odvisna od maksimalne simbolne hitrosti in od maksimalne povprečne informacije, ki jo lahko prenesemo z enim simbolom, to je s kapaciteto diskretnega kanala c_d . Velja:

$$C = 2BC_d \quad (10.32)$$

Za frekvenčno omejen analogen Gaussov kanal je potem kapaciteta dana z izrazom:

$$C = B \log_2 (1 + SNR) \quad (10.33)$$

kar je znano kot Shannon-Hartleyev teorem za kapaciteto kanala. Kapaciteta analognega frekvenčno omejenega kanala se izraža v enotah bitih na sekundo, torej pove koliko bitov informacije je mogoče po kanalu prenesti vsako sekundo.

Ko smo obravnavali modulatorske postopke, smo ugotovili, da s prenosom z več nosilci lahko boljše izkoriščamo zmogljivosti kanala, kadar šum na kanalu ni bel oziroma, kadar karakteristika kanala ni ravna. V mejnem primeru lahko kanal razdelimo na infinitezimalno majhne frekvenčne pasove širine df . V vsakem frekvenčnem pasu imamo različno razmerje med signalom in šumom $SNR(f)$. Kapaciteta frekvenčnega odseka pri frekvenci f je potem enaka:

$$C(f) = df \log_2(1 + SNR(f)) \quad (10.34)$$

Da dobimo kapaciteto celotnega kanala moramo prispevke podkanalov med seboj sešteti, oziroma integrirati čez celotno pasovno širino, od spodnje mejne frekvence f_{sp} do zgornje mejne frekvence f_{zg} :

$$C = \int_{f_{sp}}^{f_{zg}} \log_2(1 + SNR(f)) df \quad (10.35)$$

Pokazati je mogoče, da je kapaciteta pri prenosu z več nosilci po zgornjem izrazu vedno večja od kapacitete pri prenosu z enim nosilcem (10.30), razen kadar je razmerje med signalom in šumom neodvisno od frekvence in sta obe kapaciteti enaki.

10.14 Spektralna učinkovitost in Shannonova meja

Spektralna učinkovitost je pomembno merilo pri digitalnem prenosu. Spektralna učinkovitost pove, koliko bitov informacije lahko prenesemo na sekundo na vsak Hz pasovne širine in je podana z enačbo:

$$u = \frac{r_b}{B} \quad (10.36)$$

kjer je r_b bitna hitrost oziroma število bitov na sekundo in B pasovna širina, ki jo potrebujemo za prenos. Maksimalna bitna hitrost je enaka kapaciteti kanala, tako da je spektralna učinkovitost vedno enaka ali manjša od spektralne učinkovitosti, ki jo dobimo, če prenašamo s kapaciteto kanala:

$$u \leq \frac{C}{B} \quad (10.37)$$

Za Gaussov kanal torej velja:

$$u \leq \log_2 \left(1 + \frac{\sigma_x^2}{\sigma_\xi^2} \right) \quad (10.38)$$

Varianco signala σ_x^2 v zgornjem izrazu lahko izrazimo s povprečno energijo E_b , ki jo potrebujemo za prenos enega bita informacije. Velja:

$$\sigma_x^2 = E_b r_b \quad (10.39)$$

Označimo sedaj z N_0 enostransko gostoto močnostnega spektra aditivnega belega Gaussovega šuma, ki se je prištel signalu na Gaussovem kanalu. Z izrazom enostranska gostota mislimo tu

gostoto, ki jo dobimo, če seštejemo gostoto moči pri negativnih in pozitivnih frekvencah, zato je za faktor dva višja od običajne (dvostranske) gostote spektra. Varianca šuma σ_ξ^2 je potem enaka:

$$\sigma_\xi^2 = N_0 B \quad (10.40)$$

Vstavimo (10.39) in (10.40) v (10.44). Dobimo:

$$u \leq \log_2 \left(1 + \frac{E_b r_b}{N_0 B} \right) \quad (10.41)$$

Če upoštevamo še (10.36) dobimo:

$$u \leq \log_2 \left(1 + u \frac{E_b}{N_0} \right) \quad (10.42)$$

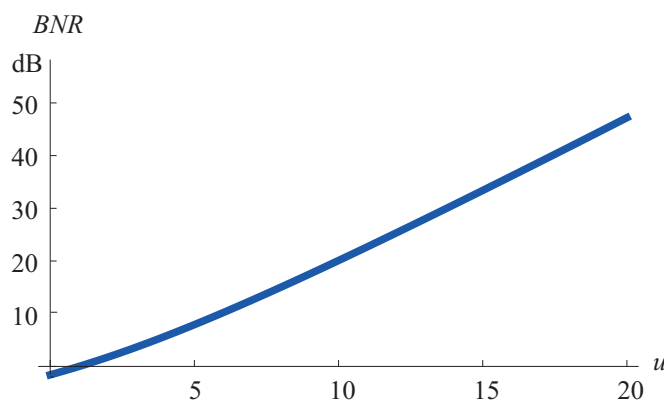
in po preureditvi:

$$\frac{E_b}{N_0} \geq \frac{2^u - 1}{u} \quad (10.43)$$

Zgornji izraz predstavlja teoretično spodnjo mejo razmerja med energijo, ki jo v povprečju potrebujemo za prenos enega bita informacije in enostransko gostoto belega Gaussovega šuma. Označimo to razmerje v decibelih z BNR (ang. Bit energy to Noise level Ratio). Velja:

$$BNR \geq 10 \log_{10} \left(\frac{2^u - 1}{u} \right) \doteq 3u - 10 \log_{10} u \quad (10.44)$$

Spodnja meja BNR v odvisnosti od spektralne učinkovitosti u je prikazana na sliki 10.5. S



Slika 10.5 – Spodnja meja razmerja med energijo bita in enosmerno gostoto šuma (BNR) v odvisnosti od spektralne učinkovitosti u . Energija, ki je potrebna za prenos enega bita informacije izredno hitro narašča s spektralno učinkovitostjo.

spektralno učinkovitostjo izredno hitro narašča energija, ki jo potrebujemo za prenos enega bita. Če želimo prenašati s spektralno učinkovitostjo 1 bit/s/Hz, je najmanjša energija, ki jo potrebujemo za prenos enega bita enaka nivoju šuma N_0 , če pa bi želeli prenašati 20 bit/s/Hz, pa bi morali biti razmerje BNR že 47,2 dB, kar pomeni, da bi morala biti energija bita že več kot 52 tisočkrat višja.

Najnižjo potrebno energijo za prenos enega bita dobimo v limiti, ko gre spektralna učinkovitost proti nič, to je ko gre potrebna pasovna širina za prenos proti neskončno:

$$BNR_{min} = \lim_{u \rightarrow \infty} \left[\log_{10} \left(\frac{2^u - 1}{u} \right) \right] = -1,59 \text{ dB} \quad (10.45)$$

Shannonova meja pove, kolikšna je pri danem šumu na Gaussovem kanalu teoretično minimalna potrebna energija za prenos enega bita informacije. Te meje ni mogoče doseči, vendar se ji sodobni kodirni postopki približajo že na 0.1 dB.

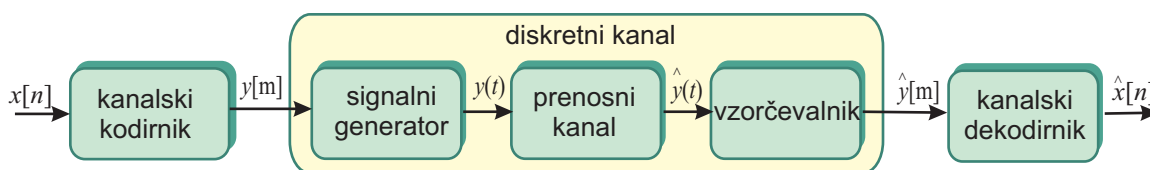
Zgornji rezultati veljajo sicer za Gaussov kanal, vendar se pogosto uporabljajo tudi za oceno spektralne učinkovitosti na drugih kanalih, kjer točna porazdelitev šuma oziroma točna karakteristika kanala ni poznana.

Kapaciteta kanala in Shannonova meja predstavlja teoretični meji spektralne učinkovitosti in energijske učinkovitosti prenosa. Tema mejama se je mogoče približati, ni pa ju mogoče doseči ali preseči. Teoretične meje so pomembne predvsem zato, da lahko dobljene rezultate primerjamo s teoretičnimi mejami. Ko so ti dovolj blizu, vemo da se ni več smiselno truditi, da bi jih še izboljšali

Kanalsko kodiranje

Poglavje obravnava različne kanalske kodirne in dekodirne postopke, od najbolj preprostih do nekaterih nekoliko zahtevnejših. Med tem, ko se preprosti kodirni postopki uporabljajo pri relativno nezahtevnem prenosu za oblikovanje spektra signala in prenos informacije o digitalnem taktu pa zahtevnejši postopki omogočajo detekcijo in/ali odpravljanje napak ter boljše izkoriščanje zmogljivosti prenosnega kanala.

V tem poglavju bomo obravnavali postopke kanalskega kodiranja in dekodiranja. Pri tem bomo predpostavili, da niz simbolov prenašamo preko diskretnega signala brez spomina in da je niz, ki ga želimo prenašati binaren niz brez redundance, oziroma, da smo redundanco že izločili z izvornim kodirnikom. Shematsko je kanalsko kodiranje in dekodiranje prikazano na sliki 11.1.



Slika 11.1 – Kanalski kodirnik preslika niz $x[n]$ na svojem vohodu v niz $y[m]$, ki je bolj prilagojen prenosu po diskretnem prenosnem kanalu. Niz na izhodu kanala $\hat{y}[m]$ kanalski dekodirnik preslika v niz $\hat{x}[n]$, za katerega želimo, da bi bil enak nizu $x[n]$

Ker je niz $x[n]$ na vohodu kanalskega kodirnika brez redundance, zanj velja, da so simboli med seboj neodvisne naključne spremenljivke z enakomerno porazdelitvijo verjetnosti. Ker je niz binaren je verjetnost enice enaka verjetnosti ničle. Entropija niza je torej enaka 1 bit/simbol. Niz $y[m]$ na izhodu kodirnika je lahko binaren ali več nivojski, zato simbolna hitrost niza na izhodu ni nujno enaka simbolni hitrosti niza na vohodu.

Pri izvornem kodiranju smo izločili redundanco, s kanalskim kodiranjem pa lahko redundanco dodajamo. Dodana redundanca je v splošnem enaka:

$$R = 1 - \frac{N}{M \log_2 L} \quad (11.1)$$

kjer je N število vhodnih bitov, ki se kodira v M L -nivojskih izhodnih simbolov. Dodana redundanca je kontrolirana. Dodajamo jo lahko za oblikovanje spektra, odkrivanje in/ali odpravljanje napak oziroma zmanjšanje verjetnosti napake BER ter za boljše izkoriščanje prenosnih zmogljivosti kanala.

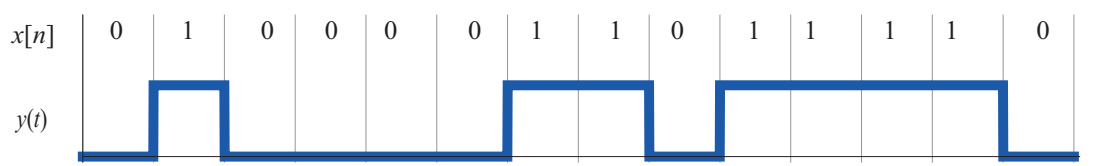
Pri dekodiranju skušamo iz sprejetega niza $\hat{y}[m]$ rekonstruirati oddani niz $x[n]$. Želimo torej,

da bi bila niza $x[n]$ in $\hat{x}[n]$ enaka. Kadar nista enaka, pravimo, da je prišlo pri prenosu do ene ali več bitnih napak.

Pri prikazu kodiranega signala bomo v nadaljevanju, zaradi večje nazornosti, predpostavili, da signalni generator generira pravokotne impulze.

11.1 Binarna koda

Ko govorimo o binarni kodi, gre pravzaprav za nekodiran prenos. Binarni niz $x[n] \in \{0, 1\}$ vodimo neposredno na vhod diskretnega kanala. Primer signala $y(t)$ pri binarni kodi je prikazan na sliki 11.2.



Slika 11.2 – Primer binarno kodiranega signala s pravokotnimi impulzi

Ker imajo lahko simboli lahko samo vrednosti 0 in 1, je njihova povprečna vrednost pri enakomerni porazdelitvi verjetnosti enako 0.5. Ker enosmerna komponenta ne vsebuje informacije, je moč, ki jo potrebujemo za njen prenos odvečna. Enosmerna komponenta v signalu moti tudi, kadar je oddajnik priklopljen na kanal preko transformatorja, preko katerega se enosmerna komponenta ne more prenašati.

Pri prenosu v višji frekvenčni legi z amplitudno modulacijo se enosmerna komponenta preslika v nosilec, kar zopet predstavlja odvečno moč pri prenosu. Zato se želimo z ustreznim kodiranjem izogniti enosmerni komponenti.

Pri sprejemu moramo vzorčenje sinhronizirati z digitalnim taktom, ki določa simbolno hitrost pri oddaji. Pri dolgih zaporedjih samih 0 ali 1 pri sprejemu izgubimo podatek za sinhronizacijo. Do tega pride lahko predvsem, kadar ima vhodni niz veliko redundanco, ki je nismo izločili z ustreznim izvornim kodiranjem. Pri digitalnem prenosu govornega signala v realnem času lahko v času tišine nastopijo dolga zaporedja samih 0.

Dekodirnik je tu preprosto odločitveno vezje oziroma komparator, ki primerja sprejete simbole $\hat{y}[n]$ z odločitvenim nivojem 0.5. Če je simbol večji od 0.5, ga prepozna za 1, če pa je manjši pa za 0.

11.2 Bipolarna koda

Z bipolarno kodo se izognemo problemu enosmerne komponente. Binarne simbole kodiramo v bipolarne po enačbi:

$$y[n] = 2x[n] - 1 \quad (11.2)$$

Na splošno je pri kodiranju je običajno preprosteje in učinkoviteje, kadar je to mogoče, uporabiti kodne tabele kot izvajati računske operacije. Za bipolarno kodo, lahko namesto zgornje enačbe

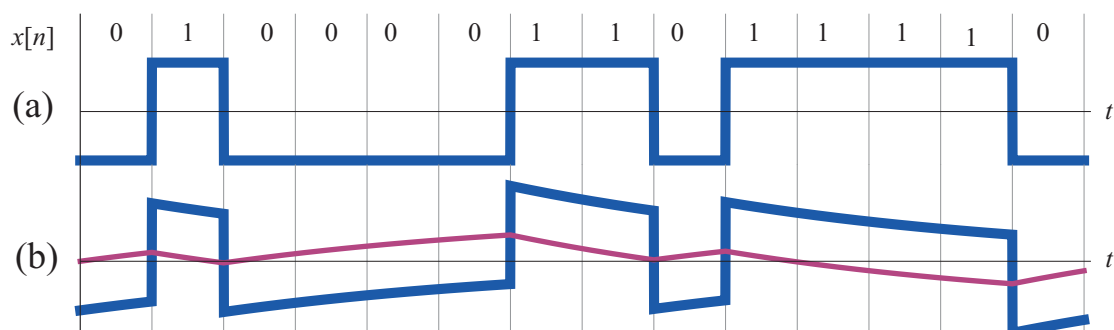
uporabimo kodno tabelo, ki je podana v tabeli 11.1.

| x | y |
|-----|-----|
| 0 | -1 |
| 1 | 1 |

Tabela 11.1 – Kodna tabela bipolarne kode

Ker so izhodni simboli dvonivojski in se en vhodni simbol preslika v enega izhodnega, bipolarna koda ne dodaja redundance.

Primer signala $y(t)$ pri binarni kodi je prikazan na sliki 11.3 a.



Slika 11.3 – Primer binarno kodiranega signala. Signal je enosmerno uravnovežen (a), zato nima enosmerne komponente. Pri prenosu preko transformatorja pa lahko pride pri daljših zaporedjih samih 0 ali 1 do lezjenja odločitvenega nivoja (b).

Z bipolarno kodo smo se sicer izognili enosmerni komponenti, še vedno pa ostaja problem sinhronizacije pri dolgih zaporedjih samih 0 ali 1. Poleg tega lahko pride pri prenosu preko transformatorja, ki ne prepušča enosmerne komponente, do lezjenja odločitvenega nivoja, na izhodu transformatorja (slika 11.3 b).

11.3 Bifazna koda

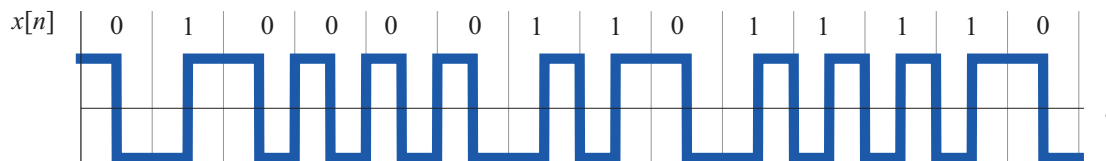
Pri bifazni kodi vsak vhodni simbol kodiramo z dvema izhodnima bipolarnima simboloma po kodni tabeli:

| x | y |
|-----|------|
| 0 | -1 1 |
| 1 | 1 -1 |

Tabela 11.2 – Kodna tabela bifazne kode

Ker se vsak vhodni bit ($N = 1$) preslika v $M = 2$ dvonivojska simbola je redundanca enaka:

$$R = 1 - \frac{N}{M} = 50\% \quad (11.3)$$



Slika 11.4 – Primer bifazne kode. Koda je enosmerno uravnorežena v okviru vsakega posameznega simbola, zato nima enosmerne komponente in tudi ne pride do problema lezenja odločitvenega nivoja

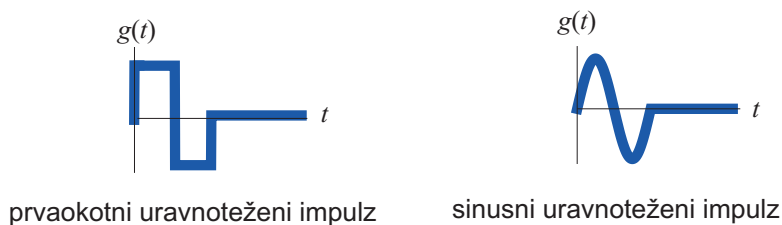
Primer kodiranega signala je prikazan na sliki 11.4.

Ker je koda uravnorežena že v okviru vsakega vhodnega simbola nima nima enosmerne komponente, gostota močnostnega spektra signala okrog frekvence pa je enaka 0. Dobro merilo, kako uspešno je izločen spekter pri nizkih frekvencah je sprotna digitalna vsota $D[n]$:

$$D[n] = \sum_{m=-\infty}^n y[m] \quad (11.4)$$

Če je absolutna vrednost $|D[n]|$ navzgor omejena, je gostota močnostnega spektra pri frekvenci 0 enaka 0. Pri bipolarni kodi je največja vrednost $|D[n]|$ enaka ena, saj se vsaki 1 prišteje -1 in obratno.

Na bifazno kodo lahko gledamo tudi kot na bifazno kodo, z uravnoreženim impulzom $g(t)$. Namesto pravokotnega impulza lahko uporabimo tudi sinusni impulz. Oba primera sta prikazana na sliki 11.5.

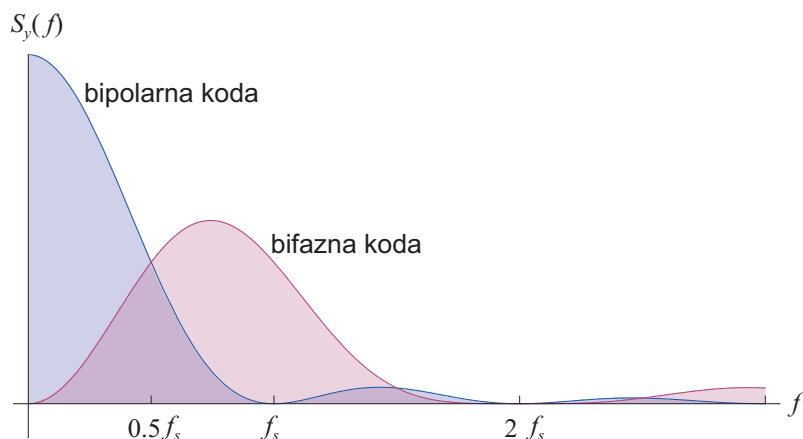


Slika 11.5 – Oblika impulzov, če na bifazno kodo gledamo kot na bipolarno kodo z uravnoreženimi impulzi

Koda se imenuje bifazna, ker je množenje uravnoreženega impulza z -1 ekvivalentno obračanju faze za 180° . Od tod izvira tudi ime bifazna koda. Ker je srednja kvadratična vrednost simbolov enaka 1 je močnostni spekter bifaznega ali bipolarnega signala v skladu z izrazom 7.20 enak:

$$S_y(f) = |G(f)|^2 \quad (11.5)$$

Primerjava močnostnega spektra bifaznega in bipolarnega signala je za pravokotne impulze podana na sliki 11.6. Vidimo, da je gostota močnostnega spektra pri bipolarni kodi različna od 0, pri bifazni kodi pa je enaka 0, zato ne pride do lezenja odločitvenega nivoja. Da bi to dosegli, je bilo potrebno dodati 50% redundance oziroma podvojiti simbolno hitrost. Zato je spekter ri bifazni kodi bistveno širši kot pri bipolarni kodi.



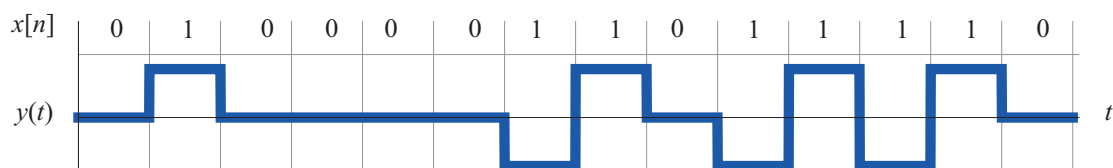
Slika 11.6 – Primerjava močnostnega spektra bifazno in bipolarno kodiranih signalov.

11.4 AMI koda

AMI (Alternate Mark Inversion) koda omejuje sprotno digitalno vsoto. Ničle na vходу se kodirajo kot ničle, enice pa izmenično kot 1 ali -1. Kodna tabela AMI kode je prikazana v tabeli 11.3. Primer kodiranega signala je prikazan na sliki 11.7. Vidimo, da so izhodni simboli

| x | y |
|-----|-----|
| 0 | 0 |
| 1 | B |

Tabela 11.3 – Kodna tabela AMI kode. Z B je označen simbol, ki ima lahko vrednost 1 ali -1. Predznak je vedno nasproten, kot pri prejšnjem simbolu B .



Slika 11.7 – Pri AMI kodi se predznak simbola 1 izmenjuje, zato je sprotna digitalna vsota omejena. Pri dolgih zaporedjih samih 0, pa lahko še vedno pride do izgube sinhronizacije.

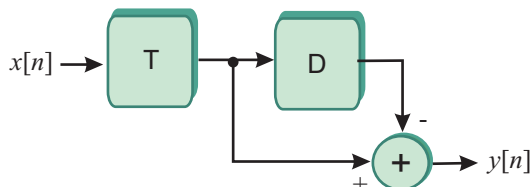
tri-nivojski ($y[n] \in \{-1, 0, 1\}$), $L = 2$). Redundanca AMI kode je torej enaka:

$$R = 1 - \frac{N}{M \log_2 L} = 36,9\% \quad (11.6)$$

Redundanca je nekoliko nižja kot pri bifazni kodi. Pri bifazni kodi je bila redundanca dodana s povečanjem simbolne hitrosti, kar je povzročilo, da je spektre signala širši. Pri AMI kodi je redundanca dodana s povečanjem števila kodnih nivojev, kar pri isti moči signala zmanjšuje razliko med dvema nivojema na polovico. Zato je povečana občutljivost na šum, zato je pri AMI

kodi za doseganje enako nizke pogostosti napake BER potrebno boljše razmerje SNR kot pri bifazni kodi.

AMI kodo lahko izvedemo s kodirnikom na sliki 11.8. Kodirnik je sestavljen iz dveh delov,



Slika 11.8 – Kodirnik AMI kode lahko naredimo iz dveh delov, T celice in linearnega kodirnika, ki je pravzaprav FIR sito z dvema koeficientoma.

predkodirnika, ki je tako imenovana T celica. T celica ob 1 na vhodu zamenja izhodno stanje ob 0 na vhodu pa pusti stanje nespremenjeno. Predkodirnik je torej neke vrste diferencialni kodirnik in ga opisuje enačba:

$$x_1[n] = x[n] \otimes x[n-1] \quad (11.7)$$

kjer je \otimes označena logična XOR operacija oziroma vsota po modulu 2. Kodirnik je linearen kodirnik, ki ga opisuje enačba:

$$y[n] = x_1[n] - x_1[n-1] \quad (11.8)$$

Močnostni spekter AMI kodiranega signala lahko sedaj izračunamo po enačbi (7.20). T celica ne vpliva na spekter, saj so simboli na izhodu T celice še vedno med seboj neodvisni z enakomerno porazdelitvijo. Kodirnik pa je pravzaprav FIR sito z dvema koeficientoma $b_0 = 1$ in $b_1 = -1$. Spekter na izhodu je tako oblikovan s prevajalno funkcijo FIR sita $H(F)$:

$$H(F) = 1 - e^{j2\pi F} \quad (11.9)$$

Na koncu lahko izrazimo spekter AMI kode kot:

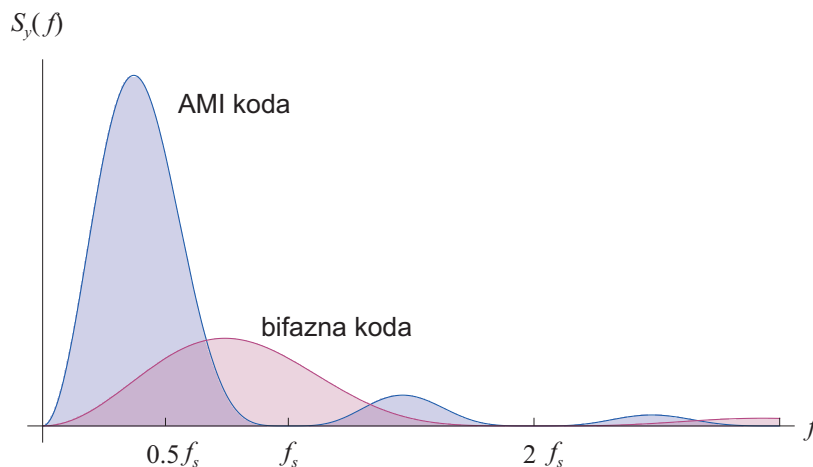
$$S_y(f) = |H(f/f_s)|^2 |G(f)|^2 \quad (11.10)$$

Primerjava spektrov AMI in bifazne kode je prikazana na sliki 11.9. Ker pri AMI kodi nismo povečali simbolne hitrosti, je osnovni snop spektra ožji kot pri bifazni kodi.

Dekodiranje ami kode je zelo preprosto, saj se vse 0 dekodirajo kot 0 vse 1 in -1 pa kod 1.

11.5 HDBn kode

HDBn kode (High Density Bipolar) z modifikacijo AMI kode. Odpravljajo problem izgube sinhronizacije pri dolgem zaporedju samih 0. Število n pove, koliko zaporednih ničel še dopuščamo v AMI kodi. Najpogosteje uporabljena je HDB3 koda, ki dovoljuje 3 zaporedne ničle. Ko se pojavijo štiri ničle pa se te nadomestijo z zaporedjem $B00V$ ali $000V$, kjer ima B enak pomen kot pri AMI kodi, V pa je enica, ki krši pravilo AMI kode in je enakega predznaka kot prejšnja enica.



Slika 11.9 – Primerjava spektrov AMI in bifazne kode. Gostota močnega spektra pri obeh kodah je pri frekvenci 0 enaka 0.

Pri dekodiranju tako zaporedje prepoznamo po simbolu V , ki krši AMI pravilo, tako, da ga lahko dekodiramo kot štiri zaporedne 0. Dve različni zaporedji uporabimo zato, da zagotovimo, da je pravilo vedno kršeno v drugo smer. Če sta bili prej dve zaporedni enici s pozitivnim predznakom, morata sedaj nastopiti dve enici z negativnim predznakom. S prvim simbolom v B v $B00V$ tako po potrebi obrnemo predznak, preden kršimo pravilo AMI kode.

11.6 Binarne blokovne kode

Binarne blokovne kode $NBMB$ so kode, ki preslikajo N vhodnih bitov kodirnika v M izhodnih bitov kodirnika, pri čemer mora biti M večji ali kvečjemu enak N . Dodana redundanca je potem enaka:

$$R = 1 - \frac{N}{M} \quad (11.11)$$

Uporablja se cela vrsta različnih binarnih blokovnih kod. Z nekaterimi se odpravlja enosmerna komponenta, druge omogočajo odkrivanje in/ali odpravljanje napak. Podrobnejša obravnava blokovnih kod presega okvire te knjige, zato si kot zgled oglejmo dve zelo preprosti kodi: 7B8B, ki omogoča odkrivanje ene napake med sedmimi bitmi in 4B64B, ki omogoča odkrivanje in popravljanje do največ 7 napak v bloku 49 bitov.

Pri obeh omenjenih kodah dodamo podatkovnim bitom tako imenovane paritetne bite. Kode, pri katerih pustimo originalne bite nespremenjene in samo dodajamo paritetne bite, imenujemo tudi sistematske kode. Blok podatkov, ki ga kodiramo imenujemo tudi kodna beseda. Z blokovno kodo preslikamo N -bitno vhodno besedo v M -bitno izhodno kodno besedo.

Pri kodi 7B8B 7 podatkovnim bitom dodamo en paritetni bit. Lahko imamo sodo ali liho pariteto. Pri sodi pariteti mora biti v kodni besedi sodo število enic pri lihi pariteti pa mora biti njihovo število liho. Paritetni bit postavimo tako, da to dosežemo.

01011101|1

00101011|0

V obeh zgornjih primerih je pariteta soda. V prvem primeru je paritetni bit enak 1, ker je število enic v vhodni besedi liho, v drugem pa je 0, saj je število enc že v vhodni besedi sodo. Če dobimo pri sprejemu liho število enic, vemo, da je prišlo pri prenosu do napake, bodisi v podatkih, bodisi v paritenem bitu.

Pri kodi 49B64B preslikamo sedem 7-bitnih vhodnih besed v osem 8-bitnih izhodnih besed. Vhodne besede razvrstimo v matriko s sedmimi vrsticami in sedmimi stolpci. Nato dodamo še paritetno vrstico in paritetni stolpec, tako da dobimo matriko dimenzij 8 x 8. V paritetnem stolpcu so paritetni biti vrstic, v paritetni vrstici pa paritetni biti stolpcev. Če nastopi napaka, je pariteta napačna v eni vrstici in v enem stolpcu, tako, da lahko natančno vemo, kje je nastopila napaka in jo lahko zato tudi popravimo, kot je to prikazano na sliki 11.10.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 |
| 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |

Slika 11.10 – Odpravljanje napake. Paritetni bit je napačen v 3 vrstici in drugem stolpcu, zato vemo, da je 0 v tretji vrstici in drugem stolpcu napačna in jo lahko popravimo na 1.

Poleg omenjenih kod obstaja cela vrsta različnih blokovnih kod. Linearne blokovne kode omogočajo, da se kodiranje izvede kot množenje vhodne kodne besede s kodirno matriko, ki vsebuje ničle in enice. Vse operacije se izvajajo po modulu 2. Pri sistematskih kodah vsebuje kodirna matrika sistematski del in paritetni del. Sistematski del je v tem primeru enotska matrika, saj zgolj preslika bite iz vhodne besede v izhodno besedo. Podobno kot kodiranje se tudi dekodiranje izvede kot množenje z matriko, pariteta pa se preverja z matriko za preverjanje paritete.

11.7 Konvolucijske kode

Za razliko od blokovnih kod se pri konvolucijskih kodah kodiranje ne izvaja po blokkih temveč sproti. Pri konvolucijskih kodah računamo izhodne bite kot linearno kombinacijo različnih vhodnih bitov, pri čemer se seštevanje izvaja po modulu.

Konvolucijski kodirnik je pomikalni register oziroma veriga zakasnilnih D celic. V vsakem časovnem koraku (digitalni uri) pride na vhod nov N novih simbolov, prejšnji simboli pa se premaknejo za N v desno. Ob vsakem časovnem koraku izračunamo tudi M izhodnih simbolov. Razmerje med številom izhodnih in številom vhodnih simbolov imenujemo tudi kodno razmerje k :

$$k = \frac{N}{M} \quad (11.12)$$

Redundanca je potem enaka:

$$R = 1 - \frac{N}{M} = 1 - k \quad (11.13)$$

Običajno se uporabljajo kode, kjer je M za ena večji od N , to je kode s kodnim razmerjem $1/2$, $2/3$, $3/4$ in tako naprej. Primer kodirnika s kodnim razmerjem $k = 1/2$ je prikazan na sliki 11.11



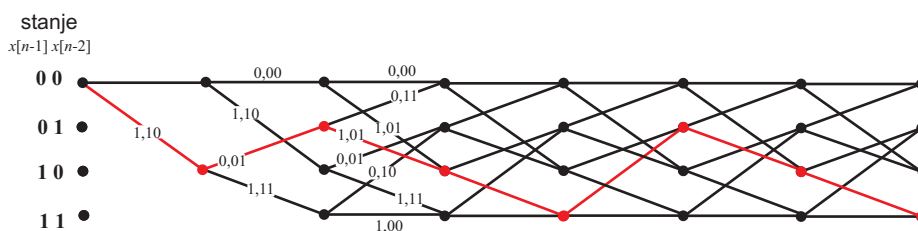
Slika 11.11 – Konvolucijski kodirnik z s kodnim razmerjem $k = 1/2$. Kodirnik je pravzaprav končni avtomat. Izhodni biti so odvisni od stanja kodirnika, to je simbolov shranjenih v D celicah in vhodnega bita. Ob vsakem urinem taktu, se spremeni stanje kodirnika glede na vhodni bit.

Stanje kodirnika je določeno s stanjem obeh D celic, ki sta lahko v stanju 0 ali 1. Ob digitalnem taktu se vhodni biti premaknejo v desno po pomikalnem registru, kar spremeni stanje kodirnika. Izhodna bita sta odvisna od stanja kodirnika (pretekli vhodni biti) in trenutnega vhodnega bita, kar opisujeta enačbi.

$$\begin{aligned} y_1[n] &= x[n] \oplus x[n-2] \\ y_2[n] &= x[n-1] \oplus x[n-2] \end{aligned} \quad (11.14)$$

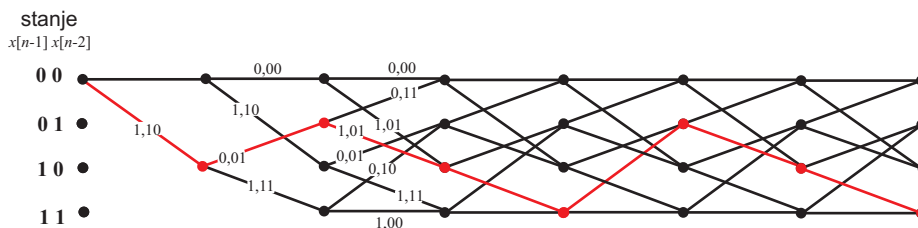
kjer je \oplus označeno seštevanje po modulu 2 oziroma operacija XOR, ki je pri binarnih simbolih ekvivalentna seštevanju po modulu 2.

Potek kodiranja je najpreprosteje prikazati v mrežnem (ang. trellis) diagramu. V diagramu je prikazan časovni potek prehodov med stanji kodirnika. Mrežni diagram za kodirnik na sliki 11.11 je prikazan na sliki 11.13.



Slika 11.12 – V mrežnem diagramu je prikazan časovni potek možnih prehodov med stanji kodirnika. Točke v mrežnem diagramu predstavljajo stanja kodirnika v določenem časovnem trenutku ob digitalnem taktu. Povezave med točkami označujejo prehode stanj. Nad povezavami so zapisani vhodni bit, ki povzroči tak prehod in izhodna dva bita, ki nastopita po prehodu. Izhodna bita sta od vhodnega ločena z vejico. Vsakemu vhodnemu sporočilu ustreza natanko ena pot skozi mrežni diagram

Vsakemu vhodnemu zaporedju pripada natanko ena pot skozi diagram. Na začetku predpostavimo, da je kodirnik v stanju 0,0. Pot, ki je v diagramu na sliki označena rdeče pripada vhodnemu zaporedju $\{1, 0, 1, 1, 0, 1, 1\}$, ustrezno izhodno zaporedje pa je $\{10, 01, 11, 01, 11\}$.



Slika 11.13 – Mrežni diagram je diagram prehodov stanj kodirnika. Veje v diagramu so označene z vhodnim bitom, ki je ločen z vejico od dveh izhodnih bitov. Veje so, za boljšo preglednost označene samo v začetnem delu diagrama. Tako same veje kot oznake se namreč ves čas ponavljajo, vse do zaključka diagrama.

11.8 Večnivojske kode

Pri več nivojskih kodah napravimo običajno samo preslikavo binarno kodiranega signala na več nivojev, pri čemer tvorimo en izhodni simbol iz več vhodnih bitov. Zato je število nivojev pri več nivojskih kodah običajno cela potenca števila dve.

Izhodi niz, ki smo ga dobili s konvolucijskim kodirnikom v prejšnjem razdelku, lahko prenašamo bipolarno, z dvakratno simbolno hitrostjo originalnega niza, ali pa več nivojsko, tako, da dva izhodna bita predstavljata en izhodni simbol iz abecede $\{3, -1, 1, 3\}$.

Pri preslikavi binarnih simbolov v večnivojske simbole je smiselno uporabiti Grayev kod, tako, da se dva sosednja nivoja razlikujeta samo v enem bitu. Napaka enega nivoja je pri sprejemu namreč ponavadi bistveno bolj verjetna kot napaka za več nivojev. Zato želimo, da pri napaki enega nivoja naredimo tudi napako samo enega bita, kar zmanjšuje BER.

Za štirinivojsko kodo je kodna tabela po Grayevem kodu prikazana v tabeli 11.4

| x | y |
|-----|-----|
| 00 | -3 |
| 01 | -1 |
| 11 | 1 |
| 10 | 3 |

Tabela 11.4 – Dvobitni Grayev kod..Dve zaporedni vhodni besedi se v Greyevem kodu razlikujeta samo v enem bitu.

Grayev kod lahko tvorimo rekurzivno. $K + 1$ bitni kod dobimo tako da kodnim besedam K bitnega koda dodamo predpono 0, njihovi zrcalni sliki pa predpono 1. Začnemo z enobitnim Garjavim kodom, ki ga predstavlja zaporedje $\{0, 1\}$. Dvobitno zaporedje potem dobimo tako da najprej besedam enobitnega zaporedja $\{0, 1\}$ dodamo predpono 0. Dobimo $\{00, 01\}$. Besedam obrnjenega (zrcalnega) zaporedja $\{1, 0\}$ dodamo predpono 1 in dobimo $\{11, 10\}$. Združeni zaporedji predstavljata dvobitni Grayev kod $\{00, 01, 11, 10\}$, kar ustreza prvemu stolpcu v tabeli 11.4.

Če postopek nadaljujemo, dobimo z združitvijo $\{000, 001, 011, 010\}$ in $\{110, 111, 101, 100\}$ tribitni Grayev kod $\{000, 001, 011, 010, 110, 111, 101, 100\}$.

11.9 Dekodiranje na osnovi največje verjetnosti

Dekodiranje na osnovi največje verjetnosti (ang. maximum likelihood – ML) je postopek, pri katerem se na sprejemni strani na osnovi sprejetih simbolov odločimo za sporočilo, ki je bilo najbolj verjetno oddano.

Da bi se pri dekodiranju res lahko odločili za sporočilo, ki je bilo oddano z največjo verjetnostjo, bi morali pri dekodiranju naenkrat dekodirati celotno sporočilo. Tak postopek pa ni praktičen in je za dolga sporočila tudi neizvedljiv. Računska zahtevnost dekodirnika pri je pri dolgih sporočilih namreč prevelika. Poleg tega bi morali čakati na sprejem celotnega sporočila, preden bi ga lahko dekodirali, kar pa, kadar gre za komunikacijo v realnem času, ni sprejemljivo.

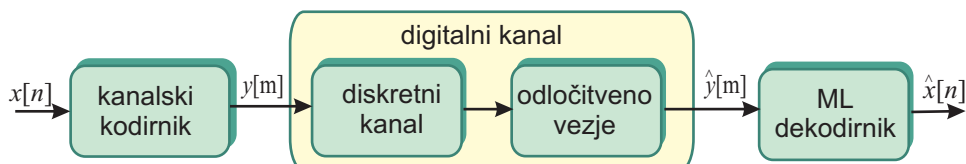
Dolžino sporočila, ki ga naenkrat dekodiramo zato omejimo na M oddanih simbolov, oziroma na N bitov vhodnega sporočila. Pri blokovnih kodah je ta dolžina določena z dolžino bloka, pri sprotnem kodiranju pa je to omejeno s kompleksnostjo dekodirnika.

Predpostavimo, da ima šum, ki se prišteje signalu, srednjo vrednost enako 0 in da verjetnost z amplitudo upada. Ta predpostavka velja za tako za Gaussov šum kot tudi za večino drugih šumov, ki so prisotni v realnih komunikacijskih kanalih. Če bi sprejemali posamezne simbole, je ob tej predpostavki najbolj verjetno, da je bil oddan simbol, ki je po vrednosti najbližje sprejetemu simbolu, ali drugače povedano, simbol ki ima od sprejetega simbola najmanjšo oddaljenost.

Vendar, kot smo že omenili, pri dekodiranju na osnovi največje verjetnosti, ne dekodiramo posameznih simbolov, temveč pri dekodiranju upoštevamo M sprejetih simbolov naenkrat, oziroma N bitov dolgo sporočilo. Pri sprejemu izračunamo oddaljenost sprejetega sporočila od vseh možnih oddanih sporočil in se nato odločimo za sporočilo, ki je od sprejetega vektorja najmanj oddaljeno. Pri tem ločimo trdo in mehko odločanje.

11.9.1 Trdo odločanje

O trdem odločanju govorimo, kadar je dekodirnik na izhodu digitalnega kanala, kar pomeni, da sta tako niz na vhodu kot niz na izhodu kanala digitalna. Digitalni kanal dobimo, če na izhod diskretnega kanala dodamo odločitveno vezje, ki se na osnovi primerjave sprejetega simbola z odločitvenimi nivoji odloči o vsakem oddanem simbolu posebej. Šele nato dodamo dekodirnik na osnovi največje verjetnosti (ML dekodirnik), kot je to prikazano na sliki 11.14.



Slika 11.14 – Pri trdem odločanju je ML dekodirnik na izhodu digitalnega kanala. Pri sprejemu se najprej odločamo o oddanih simbolih, in šele nato dekodiramo sprejeti niz $\hat{y}[m]$ na osnovi največje verjetnosti.

ML dekodirnik izračuna oddaljenost sprejetega niza $\hat{y}[m]$ od vseh možnih oddanih nizov $y[m]$ in se nato odloči, da je bil oddan niz $y[m]$, ki je od sprejetega niza najmanj oddaljen. Ta niz

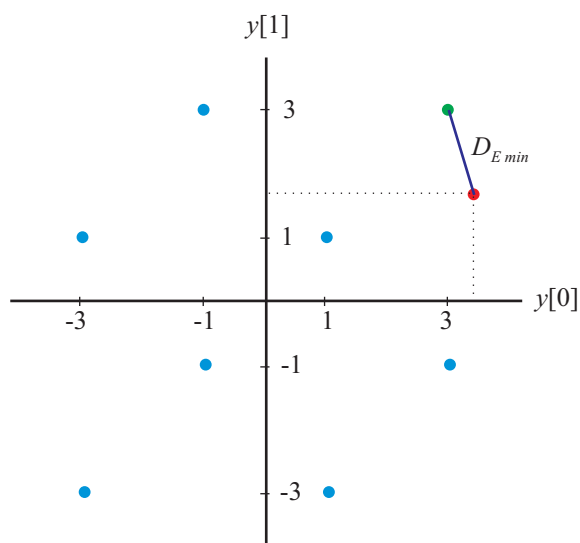
nato še dekodira v pripadajoč niz $\hat{x}[n]$. Kot merilo oddaljenosti je tu uporabljena Hammingova razdalja D_H . Hammingova razdalja je enaka vsoti absolutnih vrednosti posameznih simbolov v sporočilu:

$$D_H = \sum_{m=0}^{M-1} |y[m] - \hat{y}[m]| \quad (11.15)$$

Pri binarnih nizih je Hammingova razdalja kar enaka številu bitov v katerih se dve sporočili razlikujeta.

Če pri kodiranju ne bi dodajali redundance, bi bila možne vse kombinacije simbolov v nizu $y[n]$ in bi bil sprejeti niz vedno enak enemu od možnih oddanih nizov in z računanjem razdalje ne bi nič pridobili, saj bi bila najmanjša oddaljenost vedno enaka nič. Z dodajanjem redundance pri kodiranju dosežemo, da vsi oddani nizi niso možni, tako da sprejeti niz, kadar pride do napake, ni vedno enak enemu od oddanih nizov. Takrat se ML dekodirnik odloči za najbližji možni niz in s tem odpravi napako.

11.9.2 Mehko odločanje



Slika 11.15 – Pri dolžini sporočila $M = 2$, lahko sporočila predstavimo kot vektorje oziroma točke v ravnini. Možna sporočila so v ravni predstavljena z modrimi pikami, sprejeto sporočilo z rdečo piko, sporočilo za katero se je kodirnik odločil pa z zeleno piko. Katera sporočila so v danem trenutku možna je odvisno od kodirnega postopka. Dekodirnik se odloči za tisto med možnimi sporočili, ki je najbližje sprejetemu sporočilu. Sprejeto sporočilo je v tem primeru enako $\{3.3, 1.7\}$, zato se kodirnik odloči, da je bilo oddano sporočilo, ki je temu najbližje, to je $\{3, 3\}$

Pri trdem odločanju smo del informacije zavrgli že v odločitvenem vezju in ta ni več na voljo dekodirniku. Pri mehkem odločanju je dekodirnik neposredno na izhodu diskretnega kanala. Sprejeti simboli zato niso digitalni temveč so zvezni po amplitudi, kar pomeni da je lahko njihova vrednost poljubno realno število. Namesto Hammingove razdalje tu kot merilo uporabimo evklidsko razdaljo D_E oziroma, zaradi lažjega računanja kvadrat evklidske razdalje D_E^2 . Če M simbolov dolga sporočila predstavimo kot vektorje v M -dimenzionalnem prostoru je evklidska

razdalja kar razdalja med vektorji oziroma točkami, ki jih predstavljajo ti vektorji, torej velja:

$$D_E^2 = \sum_{m=0}^{M-1} (y[m] - \hat{y}[m])^2 \quad (11.16)$$

Če je dolžina sporočila enaka dva simbola, lahko Evklidsko razdaljo predstavimo kot razdaljo med dvema ravninskima vektorjema, kot je to prikazano na sliki 11.15.

Ker se pri mehkem odločanju ne odločimo najprej, kateri simboli so bili oddani in šele potem popravimo napako, kot pri trdem odločanju, temveč skušamo preprečiti, da bi do napake sploh prišlo, imenujemo ta postopek tudi vnaprejšnje odpravljanje napak (ang. forward error correction).

11.10 Viterbi algoritem

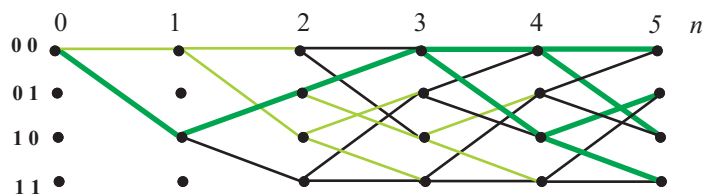
Problem pri dekodiranju na osnovi največje verjetnosti je v tem, da je potrebno izračunati oddaljenost sprejetega sporočila od vseh možnih oddanih sporočil. Pri N dolgem vhodnem sporočilu je namreč 2^N možnih sporočil, kar pomeni, da je potrebno izračunati 2^N razdalj. Ker je to eksponentna funkcija, računski zahtevnosti izredno hitro narašča, tako da mora biti N relativno majhen. Že pri $N = 10$ je potrebno izračunati 1024 razdalj, preden se lahko dekodirnik odloči o oddanem sporočilu.

Za kode, ki jih je mogoče predstaviti z mrežnim diagramom, kot so to konvolucijske kode, je mogoče z Viterbi algoritmom računsko zahtevnost dekodiranja na osnovi največje verjetnosti zmanjšati, tako da je linearno odvisna od dolžine dekodirnika N .

Kot smo že povedali pri konvolucijskih kodah, vsakemu sporočilu ustreza natanko ena pot skozi mrežni diagram. V mrežnem diagramu je 2^N različnih možnih poti, to je toliko, kot je različnih možnih sporočil. Vsaki poti skozi diagram ustreza tudi točno določeno kodirano sporočilo $y[n]$. Dekodirnik na osnovi največje verjetnosti bi moral izračunati oddaljenost sprejetega sporočila od vseh 2^N sporočil, ki pripadajo posameznim potem in se odločiti za tisto pot, ki pripada sporočilo z najmanjšo oddaljenostjo od sprejetega sporočila. Oddaljenost neke poti do sprejetega sporočila imenujemo tudi mera poti. Dekodirnik mora torej izbrati pot z najnižjo mero poti oziroma najkrajšo pot. Dolžina poti pri tem nima pomena dolžine povezav med posameznimi vozlišči temveč z njo označujemo mero poti. Za mero poti lahko uporabimo tako Hammingovo kot evklidsko razdaljo, v odvisnosti od tega ali imamo trdo ali mehko odločanje.

Osnova zmanjševanja računski zahtevnosti pri Viterbi algoritmu leži v tem, da v vsakem koraku zavržemo vse poti, ki ne morejo pripadati najkrajši poti. Najlažje si to ogledamo na primeru konvolucijskega kodirnika, ki smo ga kot primer predstavili v prejšnjem razdelku. Na sliki 11.16 je prikazan začetek njegovega mrežnega diagrama.

Z Viterbi algoritmom se pomikamo po mrežnem diagramu. V vsakem časovnem trenutku n izračunamo mere poti. V trenutku $n = 1$ sta samo dve poti, ki vodita iz stanja 00 v stanji stanji 00 in 10. Ko izračunamo njuno mero se pomaknemo v časovni trenutek $n = 2$. Tu so na voljo že štiri poti. Njihovo mero izračunamo preprosto tako, da meram poti do časovnega trenutka



Slika 11.16 – Viterbi algoritem. V vsakem vozlišču ohranimo le eno pot, ki vodi do njega, vse ostale poti zavržemo. Na ta način se število poti, ki preživijo, ne povečuje ampak je stalno enako številu stanj.

$n = 1$ prištejemo mero poti med $N = 2$ in $n = 2$. V časovnem trenutku $n = 3$ je možnih že osem poti. Mer vseh osem poti izračunamo tako, da meram do časovnega trenutka $n = 2$ prištejemo še mere med $n = 2$ in $n = 3$. Vidimo, da v vsako vozlišče tu vodita dve poti. Na sliki je so z zeleno označene poti, ki imajo do tu manjšo mero.

V časovnem trenutku $n = 3$ še ne vemo, kakšne bodo mere poti v nadaljevanju, zato tudi ne moremo vedeti skozi katero od štirih vozlišč, ki predstavljajo stanja, bo šla najkrajša pot. Sedaj vzemimo, da bo na koncu najkrajša pot tista, ki gre v času $n = 3$ skozi stanje 00. Če bo to na koncu res najkrajša pot mora biti tudi del do časovnega trenutka $n = 3$ najkrajši, torej je to pot, ki začne v stanju 00 in gre v stanja 10, 01 in pride v 00. Če bi namreč do vozlišča prišli po daljši poti 00, 00, 00, 00, bi bila namreč tudi celotna pot na koncu daljša. Ker vemo, da ta pot ne more biti del najkrajše poti skozi celoten diagram, jo lahko zavržemo in ohranimo le pot, ki je v sliki označena zeleno. Pot, ki jo ohranimo imenujemo tudi preživela pot (ang. survival path). Na podoben način lahko tudi v ostalih vozliščih ohranimo le eno preživelo pot.

Na ta način so nam namesto osmih ostale le štiri poti, oziroma osem poti, ki vodijo v vozlišča v naslednjem časovnem trenutku, kjer lahko postopek ponovimo, in zopet v vsakem vozlišču ohranimo samo eno preživelo pot.

Poti, ki so preživele do časovnega trenutka $n = 5$ so na sliki označeno s debelejšo in temnejše zeleno črto. Če se pomikamo po teh poteh nazaj, vidimo, da se vse združijo v časovnem trenutku $n = 3$. Najkrajša pot do časovnega trenutka $n = 3$ je torej zagotovo pot, ki je označena temno zeleno, zato je to tudi najbolj verjetna pot. Torej se lahko kodirnik v tem trenutku odloči, da je najbolj verjetno sporočilo, ki pripada tej poti, to je v našem primeru pot 00, 10, 01, 00.

Ker imamo omejeno dolžino kodirnika, se lahko zgodi, da se vse poti, ki so preživele nikjer ne združijo. Takrat se mora kodirnik odločiti, za najkrajšo med njimi, čeprav ni zagotovila, da je to pot, ki bo v nadaljevanju preživela. Tu se torej kodirnik lahko odloči tudi za manj verjetno pot. Daljši ko je kodirnik, manjša je verjetnost, da bo prišlo do takega primera, zato lahko z daljšim kodirnikom dobimo boljše rezultate oziroma dosežemo manjšo verjetnost napake.

Ker je računski zahtevnost samo linearno odvisna od dolžine kodirnika, je dolžina lahko mnogo večja, kot bi bila brez Viterbi algoritma, vendar je pri visokih prenosnih hitrostih še vedno zelo omejena.

Poleg Viterbi algoritma obstajajo še drugi postopki, ki omogočajo računsko bistveno učinkovitejše dekodiranje, ki je blizu dekodiranja na osnovi največje verjetnosti, vendar lahko, ker zaradi računske učinkovitosti omogočajo še bistveno daljše kodirnike, z njimi dosežemo še nižje verjet-

nosti napak kot z Viterbi algoritmom.

Pogosto se v ta namen uporabljajo tako imenovane turbo kode in kode z nizko gostoto paritetne matrike (Low Density Parity Codes – LDPC), vendar obravnava ustreznih kodirnih in dekodirnih postopkov presega okvir te knjige.

Namesto zaključka

Upam, da ste bralci, ki ste se prebili skozi snov predstavljeno v tej knjigi, to snov tudi razumeli in vam bo koristila pri vašem nadaljnjem delu.

Izredno zahtevno snov sem se potridil predstaviti na čimbolj preprost in razumljiv način. Upam, da mi je to tudi uspelo.

Ker se zavedam, da ste bralci te knjige predvsem moji študenti, vam želim veliko znanja in uspeha na izpitih.

Avtor
Sašo Tomažič